

```
RN [2] SEQUENCE FROM N.A.
RP STRAIN=RED DUROC;
RX MEDLINE=95354835; PubMed=7628604;
RA Zhao C., Ganz T., Lehrer R.I.;
RT "The structure of porcine proteogrin genes.";
RL FEBS Lett. 368:197-202(1995).
RN [3]
RP SEQUENCE OF 131-148.
RC TISSUE=Leukocyte;
RX MEDLINE=93377946; PubMed=8335113;
RA Kokryakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,
  Aleshina G.M., Shamova O.V., Korneva H.A., Lehrer R.I.;
RT "Protegrins: leukocyte antimicrobial peptides that combine features
  of corticostatic defensins and tachyplesins.";
RL FEBS Lett. 327:231-236(1993).
RN [1] FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA
CC MONOCYTOGENES AND C.ALBCANS, IN VITRO.
CC -1- SIMILARITY: BELONGS TO THE CATHHELICIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83267; CAA58240.1; -
DR EMBL; X84095; CAA58891.1; -
DR PIR; S34587; S34587.
DR HSSP; P32194; 1PG1.
DR InterPro; IPR001894; Cathelicidin.
DR Pfam; PF00666; Cathelicidins; 1.
DR ProDom; PD001838; Cathelicidin; 1.
DR PROSITE; PS00946; CATHHELICIDINS_1; 1.
DR PROSITE; PS00947; CATHHELICIDINS_2; 1.
DR Antibiotic; Signal; Amidation; Multigene family.
KW SIGNAL 1 29 POTENTIAL.
FT PROPEP 30 130
FT CHAIN 131 148
FT MOD_RES 30 30
FT PYRROLIDONE CARBOXYLIC ACID (BY
  SIMILARITY).
FT DISULFID 85 96
FT DISULFID 107 124
FT DISULFID 136 145
FT DISULFID 138 143
FT MOD_RES 148 148
FT AMIDATION (G-149 PROVIDE AMIDE GROUP).
SQ SEQUENCE 149 AA; 16578 MW; 6F4BA98429CD6ED4 CRC64;

Query Match 59.3%; Score 54; DB 1; Length 149;
Best Local Similarity 61.1%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFFSTGCR 18
  ||| | ||| |
Db 131 RGGGLCYRRRCVCVGR 148

RESULT 4
PG5_PIG 4
ID PG5_PIG STANDARD; PRT; 149 AA.
AC P49934;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEGRIN 5 PRECURSOR (PG-5).
GN NPG5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
```

```
RN [1] SEQUENCE FROM N.A.
RP STRAIN=RED DUROC;
RX MEDLINE=95354835; PubMed=7628604;
RA Zhao C., Ganz T., Lehrer R.I.;
RT "The structure of porcine proteogrin genes.";
RL FEBS Lett. 368:197-202(1995).
RN [1] FUNCTION: MICROBICIDAL ACTIVITY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CATHHELICIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X84096; CAA58892.1; -
DR HSSP; P32194; 1PG1.
DR InterPro; IPR001894; Cathelicidin.
DR Pfam; PF00666; Cathelicidins; 1.
DR ProDom; PD001838; Cathelicidin; 1.
DR PROSITE; PS00946; CATHHELICIDINS_1; 1.
DR PROSITE; PS00947; CATHHELICIDINS_2; 1.
DR Antibiotic; Amidation; Multigene family; Signal.
KW SIGNAL 1 29 POTENTIAL.
FT PROPEP 30 130
FT CHAIN 131 148
FT MOD_RES 30 30
FT PYRROLIDONE CARBOXYLIC ACID (BY
  SIMILARITY).
FT DISULFID 85 96
FT DISULFID 107 124
FT DISULFID 136 145
FT DISULFID 138 143
FT MOD_RES 148 148
FT AMIDATION (G-149 PROVIDE AMIDE GROUP)
  (BY SIMILARITY).
SQ SEQUENCE 149 AA; 16604 MW; 6CC7262429CD6864 CRC64;

Query Match 59.3%; Score 54; DB 1; Length 149;
Best Local Similarity 61.1%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFFSTGCR 18
  ||| | ||| |
Db 131 RGGGLCYRRRCVCVGR 148

RESULT 5
HSP1_MACRU
ID HSP1_MACRU STANDARD; PRT; 59 AA.
AC P42142;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Macropus rufus (Red kangaroo) (Megalela rufa).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9321;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
  SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
  SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
```

```

CC CC -!- TISSUE SPECIFICITY: TESTIS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; L35447; AAA74616.1; -
CC CC InterPro; IPR000221; Protamine_P1.
CC CC Pfam; PF00260; protamine_P1; 1.
CC CC PROSITE; PS00048; PROTAMINE_P1; 1.
CC CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
CC CC Testis; DNA condensation; Nuclear protein.
CC CC INIT_MET 0 BY SIMILARITY.
CC CC SEQUENCE 59 AA; 8230 MW; 78F1AE592B4B2FA2 CRC64;
CC CC -----
CC CC Query Match 49.5%; Score 45; DB 1; Length 59;
CC CC Best Local Similarity 69.2%; Pred. No. 0.23;
CC CC Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC CC -----
CC CC QY 1 RGGRLSYSRRRFS 13
CC CC ||| |||||:|
CC CC Db 41 RGGRRGYSRRRYS 53
CC CC -----
CC CC RESULT 6
CC CC HSP1_CAEFU STANDARD; PRT; 60 AA.
CC CC AC P42133;
CC CC DT 01-NOV-1995 (Rel. 32, Created)
CC CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
CC CC DE SPERM PROTAMINE P1.
CC CC GN PRM1.
CC CC OS Caenolestes fuliginosus.
CC CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Metatheria; Paucituberculata; Caenolestidae; Caenolestes.
CC CC OX NCBI_TaxID=37696;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC TISSUE-Sperm;
CC CC RX MEDLINE=95215351; PubMed=7700877;
CC CC RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
CC CC RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
CC CC RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC CC -!- TISSUE SPECIFICITY: TESTIS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; L35332; AAA74598.1; -
CC CC InterPro; IPR000221; Protamine_P1.
CC CC Pfam; PF00260; protamine_P1; 1.
CC CC PROSITE; PS00048; PROTAMINE_P1; 1.
CC CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
CC CC Testis; DNA condensation; Nuclear protein.
CC CC INIT_MET 0 BY SIMILARITY.
CC CC SEQUENCE 60 AA; 8514 MW; 7630E63AD33A9B05 CRC64;
CC CC -----

```

```

CC CC Query Match 49.5%; Score 45; DB 1; Length 60;
CC CC Best Local Similarity 69.2%; Pred. No. 0.24;
CC CC Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC CC -----
CC CC QY 1 RGGRLSYSRRRFS 13
CC CC ||| |||||:|
CC CC Db 41 RGGRRGYSRRRYS 53
CC CC -----
CC CC RESULT 7
CC CC HSP1_DASVI STANDARD; PRT; 60 AA.
CC CC AC P42135; P42133;
CC CC DT 01-NOV-1995 (Rel. 32, Created)
CC CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC CC DE SPERM PROTAMINE P1.
CC CC GN PRM1.
CC CC OS Dasyurus viverrinus (Southeastern quoll), and
CC CC OS Dasyurus hallucatus (Satellinus/northern quoll).
CC CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasyurus.
CC CC OX NCBI_TaxID=9279, 9280;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC TISSUE-Sperm;
CC CC RX MEDLINE=95215351; PubMed=7700877;
CC CC RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
CC CC RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
CC CC RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC CC -!- TISSUE SPECIFICITY: TESTIS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; L35340; AAA74599.1; -
CC CC DR EMBL; L35341; AAA56795.1; -
CC CC DR InterPro; IPR000221; Protamine_P1.
CC CC Pfam; PF00260; protamine_P1; 1.
CC CC PROSITE; PS00048; PROTAMINE_P1; 1.
CC CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
CC CC Testis; DNA condensation; Nuclear protein.
CC CC INIT_MET 0 BY SIMILARITY.
CC CC SEQUENCE 60 AA; 8246 MW; 615D3D85E7123025 CRC64;
CC CC -----
CC CC Query Match 49.5%; Score 45; DB 1; Length 60;
CC CC Best Local Similarity 69.2%; Pred. No. 0.24;
CC CC Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC CC -----
CC CC QY 1 RGGRLSYSRRRFS 13
CC CC ||| |||||:|
CC CC Db 42 RGGRRGYSRRRYS 54
CC CC -----
CC CC RESULT 8
CC CC HSP1_MACAG STANDARD; PRT; 60 AA.
CC CC ID HSP1_MACAG
CC CC AC P42137;
CC CC DT 01-NOV-1995 (Rel. 32, Created)
CC CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
CC CC DE SPERM PROTAMINE P1.
CC CC GN PRM1.

```



OS Macropus agilis (Agile wallaby).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.  
 OX NCBI\_TaxID=9313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPERM;  
 RA MEDLINE=95215351; PubMed=7700877;  
 RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;  
 RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
 CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; L35451; AAA74615.1; -;  
 DR InterPro; IPR000221; Protamine\_P1.  
 DR Pfam; PF00260; protamine\_P1; 1.  
 DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 FT TESTIS; DNA condensation; Nuclear protein.  
 FT INIT\_MET 0 BY SIMILARITY.  
 SQ SEQUENCE 60 AA; 8338 MW; 96255C818921EB85 CRC64;  
 -----  
 Query Match 49.58; Score 45; DB 1; Length 60;  
 Best Local Similarity 69.28; Pred. No. 0.24;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 RGGRLSYSRRRFS 13  
 ||| |||||:  
 Db 41 RGRRGYSRRRYS 53  
 -----  
 RESULT 9  
 HSP1\_ANTLA  
 ID HSP1\_ANTLA STANDARD; PRT; 61 AA.  
 AC O18745;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS Antechinus laniger.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.  
 OX NCBI\_TaxID=60701;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=97446280; PubMed=9299228;  
 RA Krajewski C., Blacket M., Buckley L., Westernman M.;  
 RT "A multigene assessment of phylogenetic relationships within the  
 RT dasyurid marsupial subfamily Sminthopsinae";  
 RL Mol. Phylogenet. Evol. 8:236-248(1997).  
 CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AF001587; AAB91377.1; -;  
 DR InterPro; IPR000221; Protamine\_P1.  
 DR Pfam; PF00260; protamine\_P1; 1.  
 DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 KW Testis; DNA condensation; Nuclear protein.  
 FT INIT\_MET 0 BY SIMILARITY.  
 SQ SEQUENCE 61 AA; 8409 MW; E021567627E562B1 CRC64;  
 -----  
 Query Match 49.58; Score 45; DB 1; Length 61;  
 Best Local Similarity 69.28; Pred. No. 0.24;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 FGGRLSYSRRRFS 13  
 ||| |||||:  
 Db 42 FGRRGYSRRRYS 54  
 -----  
 RESULT 10  
 HSP1\_ANTSW  
 ID HSP1\_ANTSW STANDARD; PRT; 61 AA.  
 AC P4213C; P42146;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS Antechinus swainsonii, Phascosorex dorsalis,  
 OS Neophascogale lorentzii (Long-clawed marsupial mouse),  
 OS Dasyurus albopunctatus (Native cat),  
 OS Dasyurus geoffroyi (Chuditch/western quoll), and  
 OS Dasyurus spartacus (Native cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.  
 OX NCBI\_TaxID=9284, 9295, 32551, 32545, 63143, 32546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=A.swainsonii, and P.dorsalis; TISSUE=SPERM;  
 RX MEDLINE=95215351; PubMed=7700877;  
 RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;  
 RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=N.lorentzii, D.albopunctatus, D.geoffroyi, and D.spartacus;  
 RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;  
 RT "Reconstructing the taxonomic radiation of dasyurine marsupials with  
 RT cytochrome b, 12S rRNA, and protamine P1 gene trees";  
 RL J. Mammal. Evol. 4:217-236(1997).  
 CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; L35338; AAB95429.1; -;  
 DR EMBL; L35339; AAA74601.1; -;  
 DR EMBL; AF010267; AAB69297.1; -;  
 DR EMBL; AF010272; AAB69302.1; -;

```

DR EMBL; AF010274; AAB69304.1; -
DR EMBL; AF010275; AAB69305.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
  BY SIMILARITY.
SQ SEQUENCE 61 AA; 8390 MW; E021472785E71221 CRC64;

Query Match          49.5%; Score 45; DB 1; Length 61;
Best Local Similarity 69.2%; Pred. No. 0.24;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRRFS 13
DB 42 RGRRGYSRRRRYS 54

RESULT 11
HSP1_PARB1
ID HSP1_PARB1 STANDARD; PRT; 61 AA.
AC O18768;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Parantechinus bilarini (Broad-footed marsupial mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Parantechinus.
OX NCBI_TaxID=32555;
RN [1]
RP SEQUENCE FROM N.A.
RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
  cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
  SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
  SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF010277; AAB69307.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
  BY SIMILARITY.
SQ SEQUENCE 61 AA; 8421 MW; C02857DF087FC9A9 CRC64;

Query Match          49.5%; Score 45; DB 1; Length 61;
Best Local Similarity 69.2%; Pred. No. 0.24;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRRFS 13
DB 43 RGRRGYSRRRRYS 55

RESULT 12
HSP1_PARB1
ID HSP1_PARB1 STANDARD; PRT; 61 AA.
AC O18768;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Parantechinus bilarini (Broad-footed marsupial mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Parantechinus.
OX NCBI_TaxID=32555;
RN [1]
RP SEQUENCE FROM N.A.
RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
  cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
  SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
  SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF010277; AAB69307.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
  BY SIMILARITY.
SQ SEQUENCE 61 AA; 8421 MW; C02857DF087FC9A9 CRC64;

Query Match          49.5%; Score 45; DB 1; Length 61;
Best Local Similarity 69.2%; Pred. No. 0.24;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRRFS 13
DB 43 RGRRGYSRRRRYS 55

RESULT 13
HSP1_DASRO
ID HSP1_DASRO STANDARD; PRT; 62 AA.
AC P42134; P42144; P42149;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Dasykaluta rosamondae, Parantechinus apicalis (Dibbler), and
  Pseudantechinus macdonnellensis (Fat-tailed marsupial mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

HSP1_SARHA
ID HSP1_SARHA STANDARD; PRT; 61 AA.
AC P42151;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Sarcophilus harrisii (Tasmanian devil), and
  Dasyurus maculatus (Tiger quoll)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sarcophilus.
OX NCBI_TaxID=9305, 9281;
RN [1]
RP SEQUENCE FROM N.A.
RA SPECIES=S.harrisii; TISSUE=Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Relief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA SPECIES=D.maculatus;
RX Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
  cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
  SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
  SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35324; AAA74608.1; -
DR EMBL; AF010276; AAB69306.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
  BY SIMILARITY.
SQ SEQUENCE 61 AA; 8410 MW; 4A215D3D85E71230 CRC64;

Query Match          49.5%; Score 45; DB 1; Length 61;
Best Local Similarity 69.2%; Pred. No. 0.24;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRRFS 13
DB 42 RGRRGYSRRRRYS 54

RESULT 13
HSP1_DASRO
ID HSP1_DASRO STANDARD; PRT; 62 AA.
AC P42134; P42144; P42149;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Dasykaluta rosamondae, Parantechinus apicalis (Dibbler), and
  Pseudantechinus macdonnellensis (Fat-tailed marsupial mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasykaluta.  
 OX NCBI\_TaxID=33560, 9291, 9299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Sperm;  
 RX MEDLINE=95215351; PubMed=7700877;  
 RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
 RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
 CC -1- FUNCTION: PROTIMES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L35325; AAA74605.1; -;  
 DR EMBL; L35326; AAA74607.1; -;  
 DR EMBL; L35337; AAA74603.1; -;  
 DR InterPro: IPR000221; Protamine.P1.  
 DR Pfam: PF00260; Protamine.P1; 1.  
 DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
 DR Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 KW Testis; DNA condensation; Nuclear protein.  
 KW INIT\_MET 0 BY SIMILARITY.  
 FT SEQUENCE 62 AA; 8595 MW; 99C02857DF087FC9 CRC64;  
 SQ

Query Match 49.5%; Score 45; DB 1; Length 62;  
 Best Local Similarity 69.2%; Pred. No. 0.24;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGRRLSYRRRFS 13  
 Db ||| |||||  
 43 RGRRRGYRRRYS 55

RESULT 14  
 HSP1\_MURLO  
 ID HSP1\_MURLO STANDARD; PRT; 62 AA.  
 AC P42140; P42150; P42154;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS Murexia longicaudata,  
 OS Phascogale tapoatafa (Common wambenger),  
 OS Smynthopsis crassicaudata (Fat-tailed dunnart),  
 OS Myrmecobius fasciatus (Numbat), and  
 OS Thylacinus cynocephalus (Tasmanian wolf).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Murexia.  
 OX NCBI\_TaxID=37736, 9293, 9301, 55782, 9275;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Sperm;  
 RX MEDLINE=95215351; PubMed=7700877;  
 RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
 RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RC SPECIES=M.fasciatus, and T.cynocephalus;  
 RX MEDLINE=97368867; PubMed=9225481;  
 RA Krajewski C., Buckley L., Westerman M.;

RT "DNA phylogeny of the marsupial wolf resolved.";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 264:911-917(1997).  
 CC -1- FUNCTION: PROTIMES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L35336; AAA74600.1; -;  
 DR EMBL; L35327; AAA74606.1; -;  
 DR EMBL; L32743; AAA99478.1; -;  
 DR EMBL; U87139; AAB91327.1; -;  
 DR EMBL; U87140; AAB91328.1; -;  
 DR InterPro: IPR000221; Protamine.P1.  
 DR Pfam; PF00260; Protamine.P1; 1.  
 DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
 DR Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 KW Testis; DNA condensation; Nuclear protein.  
 KW INIT\_MET 0 BY SIMILARITY.  
 FT SEQUENCE 62 AA; 8566 MW; 99C02857CBB73429 CRC64;  
 SQ

Query Match 49.5%; Score 45; DB 1; Length 62;  
 Best Local Similarity 69.2%; Pred. No. 0.24;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGRRLSYRRRFS 13  
 Db ||| |||||  
 43 RGRRRGYRRRYS 55

RESULT 15  
 HSP1\_ANTST  
 ID HSP1\_ANTST STANDARD; PRT; 63 AA.  
 AC P42129;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS Antechinus stuartii (Brown marsupial mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.  
 OX NCBI\_TaxID=9283;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Sperm;  
 RX MEDLINE=95215351; PubMed=7700877;  
 RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
 RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
 CC -1- FUNCTION: PROTIMES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: L35335; AAB95428.1; -;  
DR InterPro: IPR000221; Protamine\_P1.  
DR Pfam: PF00260; protamine\_P1; 1.  
DR PROSITE: PS00048; PROTAMINE\_P1; 1.  
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
Testis; DNA condensation; Nuclear protein.  
FT INIT\_MET 0  
FT 0 BY SIMILARITY.  
SQ SEQUENCE 63 AA; 8722 MW; D4FF992DAAB56D61 CRC64;

Query Match 49.5%; Score 45; DB 1; Length 63;  
Best Local Similarity 69.2%; Pred. No. 0.25;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 RGGRLSYSGRRRFS 13  
||| |||||:  
Db 44 RGGRRGYSRRRYS 56

Search completed: February 12, 2002, 12:39:52  
Job time: 805 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:40 ; Search time 232.64 Seconds  
(without alignments)  
11.317 Million cell updates/sec

Title: US-09-485-571-25  
Perfect score: 91  
Sequence: 1 RGRLSYSRRRFTSTGR 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues .  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_17.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	57.1	173	2	Q9ETA8
2	47	51.6	92	12	O38024
3	47	51.6	93	12	O73508
4	47	51.6	93	12	O73509
5	47	51.6	93	12	O73512
6	47	51.6	93	12	O73514
7	47	51.6	93	12	O73527
8	47	51.6	93	12	O73529
9	47	51.6	93	12	O73531
10	47	51.6	93	12	O93139
11	47	51.6	94	12	O73525
12	47	51.6	94	12	O86541
13	47	51.6	117	11	Q9D596
14	46	50.5	94	12	O41486
15	46	50.5	231	5	O16689
16	45	49.5	61	6	O9GLQ9
17	45	49.5	61	6	O9GLQ3
18	45	49.5	61	6	O9GLP8
19	45	49.5	61	6	O9GJQ1

20	45	49.5	62	6	Q9TUC2
21	45	49.5	62	6	Q9GLQ0
22	45	49.5	63	6	Q9TUC4
23	45	49.5	65	6	Q9GLQ2
24	44	48.4	101	12	Q9ICW8
25	44	48.4	101	12	Q9ICW2
26	44	48.4	186	2	O52604
27	44	48.4	186	2	Q9R430
28	44	48.4	293	2	Q9JUJ7
29	44	48.4	351	11	Q9CVD2
30	44	48.4	372	10	Q9FJ45
31	44	48.4	475	2	Q9RF06
32	44	48.4	475	2	Q9RSY7
33	44	48.4	532	1	Q9YF56
34	44	48.4	653	2	O30977
35	43	47.3	60	12	Q9PYU9
36	43	47.3	126	1	Q9HK12
37	43	47.3	317	8	O79545
38	42	46.2	104	12	Q9E9R1
39	42	46.2	118	1	Q9YEK7
40	42	46.2	138	12	Q90110
41	42	46.2	287	2	Q9F934
42	42	46.2	427	10	Q9SNS6
43	42	46.2	747	12	O39296
44	42	46.2	1016	5	O17484
45	42	46.2	1016	5	O17485

ALIGNMENTS

RESULT 1	
Q9ETA8	
ID Q9ETA8	PRELIMINARY; PRT; 173 AA.
AC Q9ETA8:	
DT 01-MAR-2001 (TREMBLrel. 16, Created)	
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DE HYPOTHETICAL 18.9 KDA PROTEIN.	
GN ORF58.	
OS Corynebacterium equii (Rhodococcus equi).	
OG Plasmid PREAT701 (p33701), and Plasmid virulence.	
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.	
OX NCBI_TaxID=43767;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=ATCC33701; PLASMID=PREAT701 (P33701);	
RA Takai S., Sekizaki T., Kakuda T., Nakamura M., Suzuki K., Ogino N.;	
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.	
RN [2]	
RP SEQUENCE FROM N.A.	
RC STRAIN=ATCC33701, AND 103; PLASMID=PREAT701 (P33701), AND VIRULENCE;	
RX PubMed=11083803;	
RA Takai S., Hines S.A., Sekizaki T., Nicholson V.M., Alperin D.A.,	
RA Osaki M., Takamatsu D., Nakamura M., Suzuki K., Ogino N., Kakuda T.,	
RA Dan H., Prescott J.F.;	
RT "DNA sequence and comparison of virulence plasmids from Rhodococcus	
RT equi ATCC 33701 and 103.";	
RL Infect. Immun. 68:6840-6847(2000).	
DR EMBL; AF001204; BAB16667.1; -;	
DR EMBL; AF116907; AAG21761.1; -;	
KW Hypothetical protein; Plasmid.	
SQ SEQUENCE 173 AA; 18851 MW; F18A637BCA404053 CRC64;	

Query Match 57.1%; Score 52; DB 2; Length 173;  
Best Local Similarity 52.9%; Pred. No. 0.49;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 2 GGRLSYSRRRFTSTGR 18  
|||:|:|:|:|  
DB 142 GGRISYASLYTDTGCR 158

```

RESULT 2
C38024 ID O38024 PRELIMINARY; PRT; 92 AA.
AC O38024;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 10.3 KDA PROTEIN.
OS Potato virus M.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12167;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IDAHO;
RA Cavaleer T.D., Corsini D.L., Berger P.H.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF023877; AAB81273.1; -.
DR InterPro: IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
KW Hypothetical protein.
SQ SEQUENCE 92 AA; 10319 MW; B0A9AC70B579A980 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 92;
Best Local Similarity 52.9%; Pred. No. 1.7;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRSTGTGR 18
   ||| :||| : ||
Db 40 GGRSKYARRRAIAAGR 56

RESULT 3
O73508 ID O73508 PRELIMINARY; PRT; 93 AA.
AC O73508;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 11K PROTEIN (FRAGMENT).
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ASCHERSLEBEN;
RA Matousek J., Schubert J., Dedic P.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15613; CAA75702.1; -.
DR InterPro: IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
FT NON_TER 93
SQ SEQUENCE 93 AA; 10521 MW; 2DA6E2925A8C475D CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;
Best Local Similarity 52.9%; Pred. No. 1.8;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRSTGTGR 18
   ||| :||| : ||
Db 38 GGRSTYARKRRARSIGR 54

RESULT 4
O73509 ID O73509 PRELIMINARY; PRT; 93 AA.
AC O73509;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

```

```

DE 11K PROTEIN (FRAGMENT).
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ASCHERSLEBEN;
RA Matousek J., Schubert J., Dedic P.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15615; CAA75706.1; -.
DR InterPro: IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
FT NON_TER 93
SQ SEQUENCE 93 AA; 10521 MW; 2DA6E2925A8C475D CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;
Best Local Similarity 52.9%; Pred. No. 1.8;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRSTGTGR 18
   ||| :||| : ||
Db 38 GGRSTYARKRRARSIGR 54

RESULT 5
O73512 ID O73512 PRELIMINARY; PRT; 93 AA.
AC O73512;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 11K PROTEIN (FRAGMENT).
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KARLA;
RA Matousek J., Schubert J., Dedic P.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15611; CAA75698.1; -.
DR InterPro: IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
FT NON_TER 93
SQ SEQUENCE 93 AA; 10537 MW; AC2FE2A0F98659B9 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;
Best Local Similarity 52.9%; Pred. No. 1.8;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRSTGTGR 18
   ||| :||| : ||
Db 38 GGRSTYARKRRARSIGR 54

RESULT 6
O73514 ID O73514 PRELIMINARY; PRT; 93 AA.
AC O73514;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 11K PROTEIN (FRAGMENT).
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KARLA;
RA Matousek J., Schubert J., Dedic P.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

```

--DR EMBL; Y15612; CAA75700.1; --  
 DR InterPro; IPR002568; Carla\_C4.  
 DR Pfam; PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10538 MW; 11BD9CBC997BB85 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 1.8;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRSTSTGR 18  
 ||| :|:|:| :| ||  
 Db 38 GGRSTYARKRRARSIGR 54

RESULT 7  
 O73527 ID O73527 PRELIMINARY; PRT; 93 AA.  
 AC O73527;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15609; CAA75694.1; --  
 DR InterPro; IPR002568; Carla\_C4.  
 DR Pfam; PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10333 MW; AB44F80C8018A5C9 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 1.8;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRSTSTGR 18  
 ||| :|:|:| :| ||  
 Db 38 GGRSTYARKRRARSIGR 54

RESULT 8  
 O73529 ID O73529 PRELIMINARY; PRT; 93 AA.  
 AC O73529;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15616; CAA75708.1; --  
 DR InterPro; IPR002568; Carla\_C4.  
 DR Pfam; PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10507 MW; AC2FE2A0F98645C9 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 1.8;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 GGRLSYRRRSTSTGR 18  
 ||| :|:|:| :| ||  
 Db 38 GGRSTYARKRRARSIGR 54

RESULT 9  
 O73531 ID O73531 PRELIMINARY; PRT; 93 AA.  
 AC O73531;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15610; CAA75696.1; --  
 DR InterPro; IPR002568; Carla\_C4.  
 DR Pfam; PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10333 MW; AB44F80C8018A5C9 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 1.8;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 GGRLSYRRRSTSTGR 18  
 ||| :|:|:| :| ||  
 Db 38 GGRSTYARKRRARSIGR 54

RESULT 10  
 O93139 ID O93139 PRELIMINARY; PRT; 93 AA.  
 AC O93139;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KOBRA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15614; CAA75704.1; --  
 DR InterPro; IPR002568; Carla\_C4.  
 DR Pfam; PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10565 MW; 11BD8D5D69865A75 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 1.8;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 GGRLSYRRRSTSTGR 18  
 ||| :|:~|:| :| ||  
 Db 38 GGRSTYARKRRARSIGR 54

RESULT 11  
 O73525 ID O73525 PRELIMINARY; PRT; 93 AA.  
 AC O73525;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15616; CAA75708.1; --  
 DR InterPro; IPR002568; Carla\_C4.  
 DR Pfam; PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10507 MW; AC2FE2A0F98645C9 CRC64;

Query Match 51.6%; Score 47; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 1.8;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 GGRLSYRRRSTSTGR 18  
 ||| :|:|:| :| ||  
 Db 38 GGRSTYARKRRARSIGR 54

RESULT 11  
 O73525 ID O73525 PRELIMINARY; PRT; 93 AA.  
 AC O73525;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15616; CAA75708.1; --  
 DR InterPro; IPR002568; Carla\_C4.  
 DR Pfam; PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10507 MW; AC2FE2A0F98645C9 CRC64;

```
ID 073525 PRELIMINARY; PRT; 94 AA.
AC 073525;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 11K PROTEIN.
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOBRA;
RA Matousek J., Schubert J., Dedic P.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15625; CA75721.1; -
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
SQ SEQUENCE 94 AA; 10665 MW; 5236BDFD583C830A CRC64;

Query Match 51.6%; Score 47; DB 12; Length 94;
Best Local Similarity 52.9%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFFSTGSR 18
DB 38 GGRSTYARRRRSISR 54
||| :|:|:| :||

RESULT 12
ID Q86541 PRELIMINARY; PRT; 94 AA.
AC Q86541;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 11 KDA PROTEIN.
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033173; PubMed=1413539;
RA Foster G.D., Mills P.R.;
RT "The 3'-nucleotide sequence of an ordinary strain of potato virus S.";
RL Virus Genes 6:213-220(1992).
DR EMBL; S45593; AAB23462.1; -
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
SQ SEQUENCE 94 AA; 10680 MW; 2241BD8D5D59865A CRC64;

Query Match 51.6%; Score 47; DB 12; Length 94;
Best Local Similarity 52.9%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFFSTGSR 18
DB 38 GGRSTYARRRRSISR 54
||| :|:|:| :||

RESULT 13
ID Q9D596 PRELIMINARY; PRT; 117 AA.
AC Q9D596;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 4930488L21RIK PROTEIN.
GN 4930488L21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
```

```
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Cariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015647; BAB29915.1; -
DR MGD; MGI:1923059; 4930488L21RIK.
SQ SEQUENCE 117 AA; 12868 MW; 665940B7EF991419 CRC64;

Query Match 51.6%; Score 47; DB 11; Length 117;
Best Local Similarity 69.2%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFFST 14
DB 46 GGRLSHSHQEFST 58
||| :|:|:| :||

RESULT 14
ID O41486 PRELIMINARY; PRT; 94 AA.
AC O41486;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 11K PROTEIN.
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-RB;
RA Young Y.H.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U74376; AAB65087.1; -
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
SQ SEQUENCE 94 AA; 10649 MW; C8CCDEF10F00A10A CRC64;

Query Match 50.5%; Score 46; DB 12; Length 94;
Best Local Similarity 52.9%; Pred. No. 2.6; Indels 0; Gaps 0;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFFSTGSR 18
DB 38 GGRSTYGRRRRSISR 54
||| :|:|:| :||

RESULT 15
ID O16689 PRELIMINARY; PRT; 231 AA.
```



AC 016689;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE K07E8.3 PROTEIN.  
GN K07E8.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alinscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thiertry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Jones K., Kramer J.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF016678; AAB66149.1; -;  
SQ SEQUENCE 231 AA; 26411 MW; FDF4DBE1DE511EF9 CRC64;

Query Match 50.5%; Score 46; DB 5; Length 231;  
Best Local Similarity 47.1%; Pred. No. 6.7;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 1 RGGRLSYSRREFSTSTG 17  
|||:|:|:  
Db 74 RGGRVYHDKRYPNRTG 90

Search completed: February 12, 2002, 12:38:40  
Job time: 753 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:32 ; Search time 242.57 Seconds  
(without alignments)  
5.191 Million cell updates/sec

Title: US-09-485-571-26  
Perfect score: 91  
Sequence: 1 KWSFRVSYRGISYRSR 17  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	91	100.0	17	20	AAW99413 Tachyplesin deriva
2	91	100.0	17	21	AAV93617 Peptide which may
3	88	96.7	17	20	AAW99414 Tachyplesin deriva
4	79	86.8	17	16	AAW75806 Antimicrobial tach
5	79	86.8	17	21	AAV69610 Tachyplesin analog
6	75	82.4	17	16	AAW75819 Antimicrobial tach
7	75	82.4	17	21	AAV69609 Generic tachyplesi
8	75	82.4	17	21	AAV69617 Tachyplesin analog
9	74	81.3	17	16	AAW75807 Antimicrobial tach
10	74	81.3	17	16	AAW75808 Antimicrobial tach
11	74	81.3	21	16	AAW75816 Antimicrobial tach

12	74	81.3	-	35	16	AAW75810	Antimicrobial tach
13	73	80.2	17	16	AAW75822	Antimicrobial tach	
14	71	78.0	17	10	AAW91671	New lipopolysaccha	
15	71	78.0	17	11	AAW06266	Antiviral peptide.	
16	71	78.0	17	11	AAW06861	Tachyplesin I. Li	
17	71	78.0	17	13	AAW23112	Bacterial shock tr	
18	71	78.0	17	14	AAW38489	Tachyplesin-I. Tac	
19	71	78.0	17	16	AAW75805	Tachyplesin, an an	
20	71	78.0	17	19	AAW66465	Cationic peptide t	
21	71	78.0	17	21	AAW91764	Cationic peptide t	
22	71	78.0	17	21	AAW69608	Tachyplesin (TP),	
23	71	78.0	17	21	AAW69614	Tachyplesin analog	
24	71	78.0	17	22	AAW91394	Tachykinins peptid	
25	71	78.0	39	16	AAW75817	Antimicrobial tach	
26	70	76.9	17	16	AAW75803	Antimicrobial tach	
27	68	74.7	17	11	AAW08202	Gigasins II. Tachy	
28	68	74.7	17	13	AAW23114	Bacterial shock tr	
29	68	74.7	17	14	AAW38491	Tachyplesin-III. T	
30	67	73.6	17	16	AAW75820	Antimicrobial tach	
31	67	73.6	17	16	AAW75814	Antimicrobial tach	
32	67	73.6	17	21	AAW69611	Tachyplesin analog	
33	67	73.6	17	21	AAW69612	Tachyplesin analog	
34	67	73.6	17	21	AAW69613	Tachyplesin analog	
35	67	73.6	17	21	AAW69615	Tachyplesin analog	
36	67	73.6	17	21	AAW69616	Tachyplesin analog	
37	65	71.4	17	11	AAW06862	Tachyplesin II. L	
38	65	71.4	17	13	AAW23113	Bacterial shock tr	
39	65	71.4	17	14	AAW38490	Tachyplesin-II. Ta	
40	65	71.4	17	16	AAW75811	Antimicrobial tach	
41	65	71.4	17	19	AAW66466	Cationic peptide t	
42	65	71.4	17	21	AAW91765	Cationic peptide t	
43	64	70.3	17	16	AAW75815	Antimicrobial tach	
44	64	70.3	17	16	AAW75812	Antimicrobial tach	
45	63	69.2	17	20	AAW87612	Antimicrobial pept	

ALIGNMENTS

RESULT 1  
AAW99413  
ID AAW99413 standard; peptide; 17 AA.  
XX AC  
XX AAW99413:  
DT 08-JUN-1999 (first entry)  
XX  
DE Tachyplesin derivative peptide SM1726.

XX Linear; tachyplesin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.  
XX  
XX Synthetic.  
OS  
XX  
PN WO9907728-A2.  
XX  
XX  
PD 18-FEB-1999.  
XX  
XX  
PF 06-AUG-1998; 98WO-FR01757.  
XX  
XX  
PR 12-AUG-1997; 97FR-0010297.  
XX  
XX (SYNT-) SYNT:EM SA.

XX Calas B, Chavanieu A, Grassy G, Kaczorek M;  
XX WPI; 1999-190034/16.  
XX  
XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
PT as carriers to deliver active agents into cells  
XX

PS Claim 8; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the tachyplesin family of  
CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide  
CC antibiotic family which contain a beta-sheet secondary structure linked  
CC by disulphide bridges. The new derivatives are linear and lack the  
CC disulphide bridge. The novel derivatives are used to deliver active  
CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
CC antivirals and anti-inflammatory, etc.. The derivatives are non-toxic  
CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
CC including crossing the blood-brain barrier.

XX Sequence 17 AA;

Query Match 100.0%; Score 91; DB 20; Length 17;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17

Db 1 kwsfrvsyrgisyrssr 17  
|||||

RESULT 2

AA93617  
ID AAY93617 standard; peptide; 17 AA.

AC AAY93617;

DT 25-SEP-2000 (first entry)

XX Peptide which may be linked to anticancer agents.

DE Anticancer agent; cancer cell; resistance; P-glycoprotein pump;

KW cancer.

XX Unidentified.

OS WO200032237-A1.

XX 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02939.

XX 30-NOV-1998; 98FR-0015073.

XX (SYNT-) SYNT:EM SA.

XX Tensamani J, Kaczorek M, Colin De Verdiere A;

XX WPI; 2000-412166/35.

XX New composition useful for cancer treatment and prevention, contains  
PT anticancer agent and peptide vector that transports agent into cells

XX Disclosure; Page 8; 34pp; French.

XX The specification describes a pharmaceutical composition, which  
CC comprises at least one anticancer agent associated with at least one  
CC peptide that can transport it into cancer cells and which inhibits  
CC development of resistance to the anticancer agent. By using the  
CC peptide as a vector for delivery of the anticancer agent, mechanisms  
CC that cause cancer cells to become resistant to the agent, particularly  
CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
CC produced by chemical synthesis, can be coupled easily to the agent,  
CC cross mammalian cell membranes rapidly by a passive mechanism (no  
CC receptors required), and are non-toxic and non-lytic. The compositions  
CC are used to treat cancer. The present sequence represents a peptide  
CC which may be linked to the anticancer agents of the invention.

XX

SQ Sequence 17 AA;

Query Match 100.0%; Score 91; DB 21; Length 17;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17

Db 1 kwsfrvsyrgisyrssr 17  
|||||

RESULT 3

AAW99414  
ID AAW99414 standard; peptide; 17 AA.

AC AAW99414;

XX 08-JUN-1999 (first entry)

XX Tachyplesin derivative peptide SM2307.

XX Linear; tachyplesin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

OS Synthetic.

XX WO9907728-A2.

XX 18-FEB-1999.

XX 06-AUG-1998; 98WO-FR01757.

XX 12-AUG-1997; 97FR-0010297.

XX (SYNT-) SYNT:EM SA.

XX Calas B, Chavanleu A, Grassy G, Kaczorek M;

XX WPI; 1999-190034/16.

XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
PT as carriers to deliver active agents into cells

XX Claim 8; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the tachyplesin family of  
CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide  
CC antibiotic family which contain a beta-sheet secondary structure linked  
CC by disulphide bridges. The new derivatives are linear and lack the  
CC disulphide bridge. The novel derivatives are used to deliver active  
CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
CC antivirals and anti-inflammatory, etc.. The derivatives are non-toxic  
CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
CC including crossing the blood-brain barrier.

XX Sequence 17 AA;

Query Match 96.7%; Score 88; DB 20; Length 17;

Best Local Similarity 94.1%; Pred. No. 5.3e-08;

Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17

Db 1 rwsfrvsyrgisyrssr 17  
:|||||

RESULT 4

```

AAR75806
ID AAR75806 standard; peptide; 17 AA.
XX
AC AAR75806;
XX
DT 07-FEB-1996 (first entry)
XX
DE Antimicrobial tachyplesin peptide derivative.
XX
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;
KW infection.
XX
OS Synthetic.
XX
PN WO9516776-A1.
XX
PD 22-JUN-1995.
XX
PF 19-DEC-1994; 94WO-US14619.
XX
PR 17-DEC-1993; 93US-0168809.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Putman RJ, Rao AG, Rao A;
XX
DR WPI; 1995-231570/30.
XX
PT New peptide derivs. of tachyplesin - having antimicrobial activity,
PT used against plant pathogenic fungi or human or animal infections
XX
PS Claim 1; Page 29; 45pp; English.
XX
CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin
CC (AAR75805) a small peptide isolated from Japanese horseshoe crab
CC haemocytes which has antimicrobial properties. The peptide derivatives
CC also have antimicrobial activity and can be used for killing and
CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.
CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,
CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can
CC also be used for treating and preventing infection in humans and
CC animals.
XX
SQ Sequence 17 AA;

Query Match 86.8%; Score 79; DB 16; Length 17;
Best Local Similarity 76.5%; Pred. No. 1.6e-06;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17
Db 1 kwafvayrgiayrrar 17

RESULT 5
AAY69610
ID AAY69610 standard; peptide; 17 AA.
XX
AC AAY69610;
XX
DT 08-MAY-2000 (first entry)
XX
DE Tachyplesin analogue, TPA.
XX
KW Tachyplesin analogue; generic; antimicrobial; disulphide bond;
KW antifungal; antiviral; antimicrobial; transgenic plant.
XX
OS Synthetic.
XX
PH Key Location/Qualifiers
FT Modified-site 17
FT /note= "C-terminal amide"

```

```

XX
PN US6015941-A.
XX
PD 18-JAN-2000.
XX
PF 31-OCT-1997; 97US-0962034.
XX
PR 31-OCT-1997; 97US-0962034.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Rao AG;
XX
DR WPI; 2000-126327/11.
XX
PT New tachyplesin analogs useful for controlling fungal and bacterial
PT activity in agricultural and medical applications and for controlling
PT plant viruses have four cysteine substitutions -
XX
PS Example 1; Page -; 17pp; English.
XX
CC Sequences AAY69610-Y69611 and AAY69617 represent tachyplesin analogues
CC used in an exemplification of the present invention, in which the
CC native tachyplesin cysteine residues are replaced with Ala, Leu and
CC Asp, respectively. Tachyplesin (AAY69608) is a naturally occurring
CC antimicrobial peptide which contains two disulphide bonds which help
CC to maintain its tertiary structure. The invention relates to novel
CC peptide analogues of tachyplesin (Y696912-AAY69614, AAY69616) in which
CC the cysteine residues at positions 3, 7, 12 and 16 of the native
CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,
CC valine, methionine, phenylalanine or tyrosine, the same amino acid being
CC present at all four positions. Despite being unable to form
CC intramolecular disulphide bonds, the analogues are functional as
CC antimicrobial agents. The tachyplesin analogues are useful for
CC controlling fungal and viral activity in agricultural and medical
CC applications and for controlling plant viruses. They can also be
CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,
CC soya or especially maize plants to provide resistance to pathogenic fungi
CC and viruses. Note: The present sequence is not shown in the
CC specification, but is derived from the generic tachyplesin analogue
CC sequence given in column 23.
XX
SQ Sequence 17 AA;

Query Match 86.8%; Score 79; DB 21; Length 17;
Best Local Similarity 76.5%; Pred. No. 1.6e-06;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17
Db 1 kwafvayrgiayrrar 17

RESULT 6
AAR75819
ID AAR75819 standard; peptide; 17 AA.
XX
AC AAR75819;
XX
DT 07-FEB-1996 (first entry)
XX
DE Antimicrobial tachyplesin peptide derivative.
XX
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;
KW infection.
XX
OS Synthetic.
XX
PN WO9516776-A1.
XX
PD 22-JUN-1995.
XX

```

PF 19-DEC-1994; 94WO-US14619.  
 XX  
 PR 17-DEC-1993; 93US-0168809.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Putman RJ, Rao AG, Rao A;  
 XX  
 XX WPI; 1995-231570/30.  
 DR  
 XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 PT used against plant pathogenic fungi or human or animal infections  
 PT  
 XX Claim 1; Page 35; 45pp; English.  
 PS  
 XX AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.  
 XX  
 SQ Sequence 17 AA;  
 Query Match 82.4%; Score 75; DB 16; Length 17;  
 Best Local Similarity 76.5%; Pred. No. 7e-06; Mismatches 0; Indels 0; Gaps 0;  
 Matches 13; Conservative 0;  
 QY 1 KWSFRVSYRGISYRRSR 17  
 Db 1 kwkfrvxygikyrkr 17  
 RESULT 7  
 AAY69609  
 ID AAY69609 standard; peptide; 17 AA.  
 XX  
 AC AAY69609;  
 XX  
 DT 08-MAY-2000 (first entry)  
 XX  
 DE Generic tachyplesin (TP) analogue antimicrobial peptide.  
 XX  
 KW Tachyplesin analogue; generic; antimicrobial; disulphide bond;  
 KW antifungal; antiviral; antimicrobial; transgenic plant.  
 XX  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 3 /label= Ile, Val, Met, Phe, Tyr  
 FT Misc-difference 7 /label= Ile, Val, Met, Phe, Tyr  
 FT Misc-difference 12 /label= Ile, Val, Met, Phe, Tyr  
 FT Misc-difference 16 /label= Ile, Val, Met, Phe, Tyr  
 FT /label= Ile, Val, Met, Phe, Tyr  
 FT /note= "The molecule has the same amino acid at all four  
 FT of the above positions"  
 FT Modified-site 17 /note= "C-terminal amide"  
 FT  
 FT US6015941-A.  
 XX  
 PN 18-JAN-2000.  
 XX  
 PD 31-OCT-1997; 97US-0962034.  
 XX  
 PF 31-OCT-1997; 97US-0962034.  
 XX  
 PR 31-OCT-1997; 97US-0962034.

XX (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Rao AG;  
 XX  
 XX WPI; 2000-126327/11.  
 DR  
 XX New tachyplesin analogs useful for controlling fungal and bacterial  
 XX activity in agricultural and medical applications and for controlling  
 PT plant viruses have four cysteine substitutions -  
 PT  
 XX Claim 1; Column 23; 17pp; English.  
 PS  
 XX This sequence represents a generic tachyplesin (TP) analogue which has  
 CC antimicrobial activity. Tachyplesin (AAY69608) is a naturally occurring  
 CC antimicrobial peptide which contains two disulphide bonds which help  
 CC to maintain its tertiary structure. The invention relates to novel  
 CC peptide analogues of tachyplesin (V696912-AAY69614, AAY69616) in which  
 CC the cysteine residues at positions 3, 7, 12 and 16 of the native  
 CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,  
 CC valine, methionine, phenylalanine or tyrosine, the same amino acid being  
 CC present at all four positions. Despite being unable to form  
 CC intramolecular disulphide bonds, the analogues are functional as  
 CC antimicrobial agents. The tachyplesin analogues are useful for  
 CC controlling fungal and viral activity in agricultural and medical  
 CC applications and for controlling plant viruses. They can also be  
 CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,  
 CC soya or especially maize plants to provide resistance to pathogenic fungi  
 CC and viruses.  
 XX  
 SQ Sequence 17 AA;  
 Query Match 82.4%; Score 75; DB 21; Length 17;  
 Best Local Similarity 76.5%; Pred. No. 7e-06; Mismatches 0; Indels 0; Gaps 0;  
 Matches 13; Conservative 0;  
 QY 1 KWSFRVSYRGISYRRSR 17  
 Db 1 kwkfrvxygikyrkr 17  
 RESULT 8  
 AAY69617  
 ID AAY69617 standard; peptide; 17 AA.  
 XX  
 AC AAY69617;  
 XX  
 DT 08-MAY-2000 (first entry)  
 XX  
 DE Tachyplesin analogue, TPD.  
 XX  
 KW Tachyplesin analogue; generic; antimicrobial; disulphide bond;  
 KW antifungal; antiviral; antimicrobial; transgenic plant.  
 XX  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Modified-site 17 /note= "C-terminal amide"  
 FT  
 FT US6015941-A.  
 XX  
 PN 18-JAN-2000.  
 XX  
 PD 31-OCT-1997; 97US-0962034.  
 XX  
 PF 31-OCT-1997; 97US-0962034.  
 XX  
 PR 31-OCT-1997; 97US-0962034.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Rao AG;  
 XX

DR WPI; 2000-126327/11.  
 XX New tachyplesin analogs useful for controlling fungal and bacterial  
 PT activity in agricultural and medical applications and for controlling  
 PT plant viruses have four cysteine substitutions -  
 XX Example 1: Page -: 17pp; English.  
 PS Sequences AAY69610-Y69611 and AAY69617 represent tachyplesin analogues  
 CC used in an exemplification of the present invention, in which the  
 CC native tachyplesin cysteine residues are replaced with Ala, Leu and  
 CC Asp, respectively. Tachyplesin (AAY69608) is a naturally occurring  
 CC antimicrobial peptide which contains two disulphide bonds which help  
 CC to maintain its tertiary structure. The invention relates to novel  
 CC peptide analogues of tachyplesin (Y696912-AAY69616) in which  
 CC the cysteine residues at positions 3, 7, 12 and 16 of the native  
 CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,  
 CC valine, methionine, phenylalanine or tyrosine, the same amino acid being  
 CC present at all four positions. Despite being unable to form  
 CC intramolecular disulphide bonds, the analogues are functional as  
 CC antimicrobial agents. The tachyplesin analogues are useful for  
 CC controlling fungal and viral activity in agricultural and medical  
 CC applications and for controlling plant viruses. They can also be  
 CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,  
 CC soya or especially maize plants to provide resistance to pathogenic fungi  
 CC and viruses. Note: The present sequence is not shown in the  
 CC specification, but is derived from the generic tachyplesin analogue  
 CC sequence given in column 23.  
 XX  
 SQ Sequence 17 AA;

Query Match 82.4%; Score 75; DB 21; Length 17;  
 Best Local Similarity 76.5%; Pred. No. 7e-06;  
 Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 KWSFRVSYRGISYRRSR 17  
 || ||| |||| ||| |  
 Db 1 kwfrdyrgdyrrdr 17

RESULT 9  
 AAR75807  
 ID AAR75807 standard; peptide; 17 AA.  
 AC AAR75807;  
 XX  
 XX 07-FEB-1996 (first entry)  
 DT Antimicrobial tachyplesin peptide derivative.  
 DE  
 DE Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 KW infection.  
 KW Synthetic.  
 OS  
 XX WO9516776-A1.  
 PN  
 XX 22-JUN-1995.  
 PD  
 XX 19-DEC-1994; 94WO-US14619.  
 PF  
 XX 17-DEC-1993; 93US-0168809.  
 PR  
 XX (PION-) PIONEER HI-BRED INT INC.  
 PA Putman RJ, Rao AG, Rao A;  
 XX WPI; 1995-231570/30.  
 PI  
 XX  
 XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 XX used against plant pathogenic fungi or human or animal infections  
 XX  
 PS Claim 1; Page 30; 45pp; English.  
 XX  
 CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.  
 XX  
 SQ Sequence 17 AA;

Query Match 81.3%; Score 74; DB 16; Length 17;  
 Best Local Similarity 76.5%; Pred. No. 1e-05;  
 Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KWSFRVSYRGISYRRSR 17  
 || ||| |||| ||| |  
 Db 1 kwfrvnyrgikyrrqr 17

RESULT 10  
 AAR75808  
 ID AAR75808 standard; peptide; 17 AA.  
 AC AAR75308;  
 XX  
 XX 07-FEB-1996 (first entry)  
 DT Antimicrobial tachyplesin peptide derivative.  
 DE  
 DE Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 KW infection.  
 KW Synthetic.  
 OS  
 XX WO9516776-A1.  
 PN  
 XX 22-JUN-1995.  
 PD  
 XX 19-DEC-1994; 94WO-US14619.  
 PF  
 XX 17-DEC-1993; 93US-0168809.  
 PR  
 XX (PION-) PIONEER HI-BRED INT INC.  
 PA Putman RJ, Rao AG, Rao A;  
 XX WPI; 1995-231570/30.  
 PI  
 XX  
 XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 XX used against plant pathogenic fungi or human or animal infections  
 XX  
 PS Claim 1; Page 30; 45pp; English.  
 XX  
 CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.  
 XX  
 SQ Sequence 17 AA;

Query Match 81.3%; Score 74; DB 16; Length 17;  
 Best Local Similarity 76.5%; Pred. No. 1e-05;

Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||||| |||||  
 Db 1 kwlfrvtyrgikyrrqr 17

## RESULT 11

AAR75816  
 ID AAR75816 standard; peptide: 21 AA.  
 XX AC AAR75816;  
 XX DT 07-FEB-1996 (first entry)  
 XX DE Antimicrobial tachyplesin peptide derivative.  
 KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 KW infection.  
 XX OS Synthetic.  
 XX PN W09516776-A1.  
 XX PD 22-JUN-1995.  
 XX PF 19-DEC-1994; 94WO-US14619.  
 XX PR 17-DEC-1993; 93US-0168809.  
 XX PA (PION-) PIONEER HI-BRED INT INC.  
 XX PI Putman RJ, Rao AG, Rao A;  
 XX WPI; 1995-231570/30.  
 XX DR New peptide derivs. of tachyplesin - having antimicrobial activity,  
 XX used against plant pathogenic fungi or human or animal infections  
 XX PS Claim 1; Page 34; 45pp; English.

CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.

XX SQ Sequence 21 AA;

Query Match 81.3%; Score 74; DB 16; Length 21;  
 Best Local Similarity 76.5%; Pred. No. 1.3e-05;  
 Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||||| |||||  
 Db 1 kwlfrvtyrgikyrrqr 17

## RESULT 12

AAR75810  
 ID AAR75810 standard; peptide: 35 AA.  
 XX AC AAR75810;  
 XX DT 07-FEB-1996 (first entry)  
 XX DE Antimicrobial tachyplesin peptide derivative.  
 XX PI Putman RJ, Rao AG, Rao A;

KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 KW infection.  
 XX OS Synthetic.  
 XX PN W09516776-A1.  
 XX PD 22-JUN-1995.  
 XX PF 19-DEC-1994; 94WO-US14619.  
 XX PR 17-DEC-1993; 93US-0168809.  
 XX PA (PION-) PIONEER HI-BRED INT INC.  
 XX PI Putman RJ, Rao AG, Rao A;  
 XX WPI; 1995-231570/30.  
 XX DR New peptide derivs. of tachyplesin - having antimicrobial activity,  
 XX used against plant pathogenic fungi or human or animal infections  
 XX PS Claim 1; Page 31; 45pp; English.

CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.

XX SQ Sequence 35 AA;

Query Match 81.3%; Score 74; DB 16; Length 35;  
 Best Local Similarity 76.5%; Pred. No. 2.2e-05;  
 Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||||| |||||  
 Db 1 kwlfrvtyrgikyrrqr 17

## RESULT 13

AAR75822  
 ID AAR75822 standard; peptide: 17 AA.

XX AC AAR75822;

XX DT 07-FEB-1996 (first entry)

XX DE Antimicrobial tachyplesin peptide derivative.

XX KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 XX KW infection.

XX OS Synthetic.

XX PN W09516776-A1.

XX PD 22-JUN-1995.

XX PF 19-DEC-1994; 94WO-US14619.

XX PR 17-DEC-1993; 93US-0168809.

XX PA (PION-) PIONEER HI-BRED INT INC.

XX PI Putman RJ, Rao AG, Rao A;

XX



DR WPI; 1995-231570/30.

XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 PT used against plant pathogenic fungi or human or animal infections

XX Claim 1; Page 36; 45pp; English.

XX AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.

XX Sequence 17 AA;

Query Match 80.2%; Score 73; DB 16; Length 17;  
 Best Local Similarity 76.3%; Pred. No. 1.5e-05;  
 Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
 II III IIII IIII  
 Db 1 kwfrvryrgleyrrer 17

RESULT 14

AAP91671

ID AAP91671 standard; peptide; 17 AA.

XX AAP91671;

XX 29-JUN-1990 (first entry)

XX New lipopolysaccharide-binding polypeptide(s).

XX Lipopolysaccharide-binding polypeptide; bacterial infections;  
 KW lipopolysaccharide (LPS) endotoxins; antibacterial agents;  
 KW LPS-mediated immune disorders; inflammatory disorders;  
 KW horseshoe crab haemocytes.

XX Horseshoe crab.

XX Key Location/Qualifiers

FT Misc-difference 1 /label-OTHER

FT /note="H-Lys"

FT Disulfide-bond 3..16

FT Disulfide-bond 7..12

FT Misc-difference 17 /label-OTHER

FT /note="Arg-OH or Arg-NH2"

XX WO8901492-A.

XX 23-FEB-1989.

XX 19-AUG-1988; 88WO-JP00823.

XX 21-AUG-1987; 87JP-0206258.

XX (SEK) SEIKAGAKU KOGYO KK.

XX Nakamura T, Iwanaga S, Ohno M, Miyazaki K;  
 PI WPI; 1989-068854/09.

XX New lipo:polysaccharide- binding polypeptide(s) -  
 PT useful for treating bacterial infections and immune and  
 PT inflammatory disorders.

XX Claim 2; Page 27; 39pp; English.

XX The lipopolysaccharide-binding polypeptides may be prepd. by either  
 CC solid-phase peptide synthesis followed by oxidn. to form the disulphide  
 CC bridges or, subjecting horseshoe crab haemocytes to hypotonic extraction,  
 CC extracting the residue with acid, and purifying the extract. The  
 CC polypeptides have high affinity for lipopolysaccharide (LPS) endotoxins  
 CC and are useful for removing such toxins from fluids, as antibacterial  
 CC agents, eg active against Salmonella spp. and S. aureus, and for  
 CC treating LPS-mediated immune and inflammatory disorders, eg superior  
 CC tracheobronchial infections, urinary tract infections, bedsores, burns,  
 CC colitis, cirrhosis, hepatic insufficiency and post-operative  
 CC complications.

XX Sequence 17 AA;

Query Match 78.0%; Score 71; DB 10; Length 17;  
 Best Local Similarity 76.5%; Pred. No. 3.2e-05;  
 Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
 II III IIII IIII  
 Db 1 kwcfvrcyrgicyrrcr 17

RESULT 15

AAR06266

ID AAR06266 standard; peptide; 17 AA.

XX AAR06266;

XX 13-DEC-1990 (first entry)

XX Antiviral peptide.

XX Vesicular stomatitis virus; HIV; AIDS;

XX Tachypeus tridentatus.

XX Key Location/Qualifiers

FT Disulfide-bond 3..16

FT Disulfide-bond 7..12

XX JP02167230-A.

XX 27-JUN-1990.

XX 30-JUN-1989; 89JP-0166811.

XX 26-SEP-1988; 88JP-0239051.

XX 30-JUN-1989; 89JP-0166811.

XX (SEK) SEIKAGAKU KOGYO KK.

XX WPI; 1990-241996/32.

XX Antivirus agents of polypeptide - useful as antiviral agents for  
 PT vesicular stomatitis virus or human immuno-deficiency virus

XX Claim 1; Page 309; 12pp; Japanese.

XX Sequence 17 AA;

Query Match 78.0%; Score 71; DB 11; Length 17;  
 Best Local Similarity 76.5%; Pred. No. 3.2e-05;  
 Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
 II III IIII IIII  
 Db 1 kwcfvrcyrgicyrrcr 17

Search completed: February 12, 2002, 12:30:33  
Job time: 366 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:23 ; Search time 106.12 Seconds  
(without alignments)  
3.605 Million cell updates/sec

Title: US-09-485-571-26

Perfect score: 91  
Sequence: 1 KWSFRVSYRGISYRRSR 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	79	86.8	17	1	US-08-168-809-5
2	75	82.4	17	1	US-08-168-809-18
3	75	82.4	17	3	US-08-962-034-2
4	74	81.3	17	1	US-08-168-809-6
5	74	81.3	17	1	US-08-168-809-7
6	74	81.3	21	1	US-08-168-809-15
7	74	81.3	35	1	US-08-168-809-9
8	73	80.2	17	1	US-08-168-809-21
9	71	78.0	17	1	US-07-926-965-1
10	71	78.0	17	1	US-07-876-883-1
11	71	78.0	17	1	US-08-168-809-4
12	71	78.0	17	1	US-08-426-550-1
13	71	78.0	17	3	US-08-962-034-1
14	71	78.0	39	1	US-08-168-809-16
15	70	76.9	17	1	US-08-168-809-2
16	68	74.7	17	1	US-07-876-883-3
17	68	74.7	17	1	US-08-426-550-3
18	67	73.6	17	1	US-08-168-809-13
19	67	73.6	17	1	US-08-168-809-19
20	65	71.4	17	1	US-07-876-883-2
21	65	71.4	17	1	US-08-168-809-10
22	65	71.4	17	1	US-08-426-550-2
23	64	70.3	17	1	US-08-168-809-11
24	64	70.3	17	1	US-08-168-809-14
25	63	69.2	17	4	US-09-230-180-36
26	61	67.0	17	1	US-08-168-809-12
27	61	67.0	18	1	US-07-876-883-4

28	61	67.0	18	1	US-08-282-030-7	Sequence 7, Appli
29	61	67.0	18	1	US-08-426-550-4	Sequence 4, Appli
30	61	67.0	18	5	PCT-US95-10219-7	Sequence 7, Appli
31	61	67.0	19	1	US-08-282-030-8	Sequence 8, Appli
32	61	67.0	19	5	PCT-US95-10219-8	Sequence 8, Appli
33	59.5	65.4	29	1	US-08-168-809-8	Sequence 8, Appli
34	59	64.8	17	1	US-07-856-026B-13	Sequence 13, Appl
35	59	64.8	17	1	US-08-168-809-17	Sequence 17, Appl
36	59	64.8	18	1	US-07-856-026B-14	Sequence 14, Appl
37	58	63.7	18	1	US-08-037-777A-1	Sequence 1, Appli
38	58	63.7	18	1	US-07-876-883-5	Sequence 5, Appli
39	58	63.7	18	1	US-07-856-026B-23	Sequence 23, Appl
40	58	63.7	18	1	US-08-379-039C-1	Sequence 1, Appli
41	58	63.7	18	1	US-08-426-550-5	Sequence 5, Appli
42	58	63.7	18	2	US-08-459-400-1	Sequence 1, Appli
43	57	62.6	16	1	US-07-856-026B-12	Sequence 12, Appl
44	56	61.5	17	1	US-07-856-026B-3	Sequence 3, Appli
45	56	61.5	18	1	US-07-856-026B-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-08-168-809-5  
; Sequence 5, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,809  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roth, Michael J.  
; REGISTRATION NUMBER: 29,342  
; REFERENCE/DOCKET NUMBER: 0173R US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 515-245-3595  
; TELEFAX: 515-245-3634  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-168-809-5

Query Match 86.8%; Score 79; DB 1; Length 17;  
Best Local Similarity 76.5%; Pred No. 4.7e-07;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17



; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-168-809-6

Query Match 81.3%; Score 74; DB 1; Length 17;

Best Local Similarity 76.5%; Pred. No. 3.1e-06;  
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
|||:||||| |||  
Db 1 KWLFRVNYRGIKYRRQR 17

## RESULT 5

US-08-168-809-7

; Sequence 7, Application US/08168809  
; Patent No. 5580852

; GENERAL INFORMATION:

; APPLICANT: Putnam, Rebecca J.

; APPLICANT: Rao, Aragula G.

; TITLE OF INVENTION: DERIVATIVES OF TACHYPLEPSIN HAVING

; INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pioneer Hi-Bred International

; STREET: 700 Capital Square, 400 Locust Stree

; CITY: Des Moines

; STATE: IA

; COUNTRY: USA

; ZIP: 50309

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/168,809

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Roth, Michael J.

; REGISTRATION NUMBER: 29,342

; REFERENCE/DOCKET NUMBER: 0173R US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 515-245-3595

; TELEFAX: 515-245-3634

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 17 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-168-809-7

Query Match 81.3%; Score 74; DB 1; Length 17;

Best Local Similarity 76.5%; Pred. No. 3.1e-06;  
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
|||:||||| |||  
Db 1 KWLFRVNYRGIKYRRQR 17

## RESULT 6

US-08-168-809-15

; Sequence 15, Application US/08168809  
; Patent No. 5580852

; GENERAL INFORMATION:

; APPLICANT: Putnam, Rebecca J.

; APPLICANT: Rao, Aragula G.

; TITLE OF INVENTION: DERIVATIVES OF TACHYPLEPSIN HAVING

; INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pioneer Hi-Bred International

; STREET: 700 Capital Square, 400 Locust Stree

; CITY: Des Moines

; STATE: IA

; COUNTRY: USA

; ZIP: 50309

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/168,809

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Roth, Michael J.

; REGISTRATION NUMBER: 29,342

; REFERENCE/DOCKET NUMBER: 0173R US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 515-245-3595

; TELEFAX: 515-245-3634

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 21 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-168-809-15

Query Match 81.3%; Score 74; DB 1; Length 21;

Best Local Similarity 76.5%; Pred. No. 3.9e-06;  
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
|||:||||| |||  
Db 1 KWLFRVNYRGIKYRRQR 17

## RESULT 7

US-08-168-809-9

; Sequence 9, Application US/08168809

; Patent No. 5580852

; GENERAL INFORMATION:

; APPLICANT: Putnam, Rebecca J.

; APPLICANT: Rao, Aragula G.

; TITLE OF INVENTION: DERIVATIVES OF TACHYPLEPSIN HAVING

; INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pioneer Hi-Bred International

; STREET: 700 Capital Square, 400 Locust Stree

; CITY: Des Moines

; STATE: IA

; COUNTRY: USA

; ZIP: 50309

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/168,809
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Roth, Michael J.
; REGISTRATION NUMBER: 29,342
; REFERENCE/DOCKET NUMBER: 0173R US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-245-3595
; TELEFAX: 515-245-3634
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; OS-08-168-809-9

```

## RESULT

; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 US-08-168-809-21

## RESULT

```

US-07-926-965-1
: Sequence 1, Application US/07926965
: Patent No. 5416194
: GENERAL INFORMATION:
: APPLICANT: NAKAMURA, TAKANORI; IWANAGA, SADAOKI;
: APPLICANT: OHNO, MOTONORI; MIYAZAKI, KYOSUKE
: TITLE OF INVENTION: NOVEL POLYPEPTIDE AND METHOD FOR
: TITLE OF INVENTION: PREPARING THE SAME
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BIERMAN & MUSERLIAN
: STREET: 600 THIRD AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10016
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/926,965
: FILING DATE: 19920807
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/665,819
: FILING DATE: 07-MAR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/348,487
: FILING DATE: 19-APR-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: CHARLES A. MUSERLIAN
: REGISTRATION NUMBER: 19,683
: REFERENCE/DOCKET NUMBER: TSU-4B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-661-8000
: TELEFAX: 212-661-8002
: INFORMATION FOR SEQ. ID NO. 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 17 AMINO ACIDS
: TYPE: AMINO ACID
: STRANDEDNESS: SINGLE
: TOPOLOGY:
: MOLECULE TYPE: POLYPEPTIDE
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: HORSESHOE CRAB
: STRAIN: TACHYPLEUS TRIDENTATUS
: INDIVIDUAL ISOLATE:
: DEVELOPMENTAL STAGE:
: HAPLOTYPE:
: TISSUE TYPE:
: CELL TYPE: HEMOCYTE
: CELL LINE:
: ORGANELLE:
: FEATURE:
: NAME/KEY: LPS-binding polypeptide, or LPB

```



```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426.550
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Miller, Charles E.
; REGISTRATION NUMBER: 24,576
; REFERENCE/DOCKET NUMBER: 7568-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-9741
; INVENTION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-426-550-1

Query Match      78.0%; Score 71; DB 1; Length 17;
Best Local Similarity 76.5%; Pred. No. 9.7e-06;
Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17
   || ||| ||| ||| |||
Db 1 KWCFRVCYRGICYYRRCR 17

RESULT 13
US-08-962-034-1
; Sequence 1, Application US/08962034
; Patent No. 6015941
; GENERAL INFORMATION:
; APPLICANT: Rao, A. Gururaj
; TITLE OF INVENTION: PEPTIDE DERIVATIVES OF TACHYPLESIN
; TITLE OF INVENTION: HAVING ANTIMICROBIAL ACTIVITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer IP Group of Alston & Bird
; STREET: 3605 Glenwood Ave.
; CITY: Raleigh
; STATE: NC
; COUNTRY: US
; ZIP: 27622-1107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/962.034
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5718-19
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-962-034-1

Query Match      78.0%; Score 71; DB 3; Length 17;
Best Local Similarity 76.5%; Pred. No. 9.7e-06;
Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17
   || ||| ||| ||| |||
Db 1 KWCFRVCYRGICYYRRCR 17

RESULT 14
US-08-168-809-16
; Sequence 16, Application US/08168809
; Patent No. 5580852
; GENERAL INFORMATION:
; APPLICANT: Putnam, Rebecca J.
; APPLICANT: Rao, Aragula G.
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pioneer Hi-Bred International
; STREET: 700 Capital Square, 400 Locust Stree
; CITY: Des Moines
; STATE: IA
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/168.809
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Roth, Michael J.
; REGISTRATION NUMBER: 29,342
; REFERENCE/DOCKET NUMBER: 0173R US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-245-3595
; TELEFAX: 515-245-3634
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-168-809-16

Query Match      78.0%; Score 71; DB 1; Length 39;
Best Local Similarity 76.5%; Pred. No. 2.4e-05;
Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17
   || ||| ||| ||| |||
Db 1 KWCFRVCYRGICYYRRCR 17

RESULT 15
US-08-168-809-2
; Sequence 2, Application US/08168809
; Patent No. 5580852
; GENERAL INFORMATION:
```



APPLICANT: Putnam, Rebecca J.  
APPLICANT: Rao, Araguia G.  
TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pioneer Hi-Bred International  
STREET: 700 Capital Square, 400 Locust Stree  
CITY: Des Moines  
STATE: IA  
COUNTRY: USA  
ZIP: 50309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/168,809  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Roth, Michael J.  
REGISTRATION NUMBER: 29,342  
REFERENCE/DOCKET NUMBER: 0173R US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 515-245-3595  
TELEFAX: 515-245-3634  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-168-809-2

Query Match 76.9%; Score 70; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 1.4e-05;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
|||:|||||  
Db 1 KWLFRVNERGIKYRQR 17

Search completed: February 12, 2002, 12:32:24  
Job time: 452 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:40 ; Search time 126.85 Seconds  
(without alignments)  
10.209 Million cell updates/sec

Title: US-09-485-571-26

Perfect score: 91

Sequence: 1 KWSFRVSYRGISYRRSR 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	71	78.0	17	2 A38824	tachyplesin I - ho
2	71	78.0	19	2 JX0124	tachyplesin I prec
3	71	78.0	77	2 A38345	tachyplesin I prec
4	68	74.7	17	2 JX0125	tachyplesin III -
5	65	71.4	77	2 B38345	tachyplesin II pre
6	61	67.0	18	2 JU0124	polyphemus I - A
7	58	63.7	18	2 JU0125	polyphemus II -
8	47	51.6	615	2 D86473	hypothetical prote
9	44.5	48.9	88	2 C82472	conserved hypotet
10	44	48.4	307	2 F84162	hypothetical prote
11	44	48.4	378	2 A84161	hypothetical prote
12	44	48.4	620	2 E86468	protein F12K21.26
13	42	46.2	480	2 G75313	molybdate metaboli
14	42	46.2	536	2 D42463	hypothetical prote
15	41	45.1	279	2 T02495	hypothetical prote
16	41	45.1	770	2 T22808	hypothetical prote
17	40	44.0	330	2 A72534	hypothetical prote
18	40	44.0	473	1 ERAD40	early E2A DNA-bind
19	40	44.0	474	1 ERAD41	outer membrane ush
20	40	44.0	901	2 H82850	protein tyrosine k
21	40	44.0	1099	2 S48053	protein tyrosine k
22	40	44.0	1100	2 S43677	L-JAK protein-tyro
23	40	44.0	1124	2 A55747	protein-tyrosine k
24	40	44.0	1299	2 F58401	hypothetical prote
25	39.5	43.4	781	2 A86205	hypothetical prote
26	39.5	43.4	846	2 T19179	probable high-affi
27	39	42.9	242	2 D72485	probable oligopept
28	39	42.9	597	2 B71130	protein F1504.37 [
29	39	42.9	767	2 G86476	

30	39	42.9	968	2 S46992	protein pl30 - rat
31	39	42.9	1148	2 F86403	probable transposo
32	38.5	42.3	563	2 B70918	hypothetical prote
33	38	41.8	94	2 I59528	MHC HLA-DQ-beta ce
34	38	41.8	203	2 C25511	cc protein - fruit
35	38	41.8	223	2 S68196	hypothetical prote
36	38	41.8	261	1 HLH02C	MHC class II histo
37	38	41.8	327	2 A70860	hypothetical prote
38	38	41.8	329	1 D71316	conserved hypotet
39	38	41.8	330	2 E96503	protein F9C16.11 [
40	38	41.8	358	2 H75264	hypothetical prote
41	38	41.8	360	2 T20886	hypothetical prote
42	38	41.8	411	2 S46800	LAG1 protein - yea
43	38	41.8	466	2 E75201	pyridoxal phosphat
44	38	41.8	484	1 ERAD12	early E2A DNA-bind
45	38	41.8	514	2 B72752	hypothetical prote

ALIGNMENTS

RESULT 1

A38824

tachyplesin I - horseshoe crab (Tachyplesus gigas)

C:Species: Tachyplesus gigas

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997

C:Accession: A38824

R:Muta, T.: Fujimoto, T.; Nakajima, H.; Iwanaga, S.

J. Biochem. 108, 261-266, 1990

A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (C

ssing intermediate of its precursor.

A:Reference number: JX0124; MUID:91035357

A:Accession: A38824

A:Molecule type: protein

A:Residues: 1-17 <MUT>

A:Experimental source: hemocyte

C:Keywords: amidated carboxyl end

F;3-16,7-12/Disulfide bonds: #status predicted

F;17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match

Best Local Similarity 78.0%; Score 71; DB 2; Length 17;

Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17

DB 1 KWCPRVCYRGICYRRCR 17

RESULT 2

JX0124

tachyplesin I precursor - horseshoe crab (Carcinoscorpius rotundicauda)

C:Species: Carcinoscorpius rotundicauda

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997

C:Accession: JX0124

R:Muta, T.: Fujimoto, T.; Nakajima, H.; Iwanaga, S.

J. Biochem. 108, 261-266, 1990

A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (C

ssing intermediate of its precursor.

A:Reference number: JX0124; MUID:91035357

A:Accession: JX0124

A:Molecule type: protein

A:Residues: 1-19 <MUT>

A:Experimental source: hemocyte

C:Keywords: amidated carboxyl end

F;1-17/Product: tachyplesin I #status experimental <MAT>

F;3-16,7-12/Disulfide bonds: #status predicted

F;17/Modified site: amidated carboxyl end (Arg) (amide in mature form from followi

Query Match

Best Local Similarity 78.0%; Score 71; DB 2; Length 19;

Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||| ||| ||| ||| |||

Db 1 KWCFRVCYRGICYRRCR 17

# RESULT 3

A38345  
 tachyplesin I precursor - horseshoe crab (Tachyplesus tridentatus)

C:Species: Tachyplesus tridentatus  
 C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 01-Dec-2000

C:Accession: A38345; A30068

R:Shigenaga, T.; Muta, T.; Toh, Y.; Tokunaga, F.; Iwanaga, S.

J. Biol. Chem. 265, 21350-21354, 1990

A:Title: Antimicrobial tachyplesin peptide precursor. cDNA cloning and cellular localization

A:Reference number: A38345; MUID:91065956

A:Accession: A38345

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-77 <SHI>

C:Cross-references: GB:M57242; GB:J05689; NID:g161659; PID:g161660

R:Nakamura, T.; Furunaka, H.; Miyata, T.; Tokunaga, F.; Muta, T.; Iwanaga, S.; Niwa, M.

J. Biol. Chem. 263, 16709-16713, 1988

A:Title: Tachyplesin, a class of antimicrobial peptide from the hemocytes of the horseshoe crab

A:Reference number: A30068; MUID:89034158

A:Accession: A30068

A:Molecule type: protein

A:Residues: 24-40 <NAK>

Query Match 78.0%; Score 71; DB 2; Length 77;

Best Local Similarity 76.5%; Pred. No. 7.1e-05;

Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||| ||| ||| ||| |||

Db 24 KWCFRVCYRGICYRRCR 40

# RESULT 4

JX0125

tachyplesin III - horseshoe crab (Tachyplesus gigas)

C:Species: Tachyplesus gigas

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997

C:Accession: JX0125

R:Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.

J. Biochem. 108, 261-266, 1990

A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (Caridea)

A:Reference number: JX0124; MUID:91035357

A:Accession: JX0125

A:Molecule type: protein

A:Residues: 1-17 <MUT>

A:Experimental source: hemocyte

C:Keywords: amidated carboxyl end

F:3-16,7-12/Disulfide bonds: #status predicted

F:17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 74.7%; Score 68; DB 2; Length 17;

Best Local Similarity 70.6%; Pred. No. 4.9e-05;

Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||| ||| ||| |||

Db 1 KWCFRVCYRGICYRRCR 17

# RESULT 5

B38345

tachyplesin II precursor - horseshoe crab (Tachyplesus tridentatus)

C:Species: Tachyplesus tridentatus

C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 08-Dec-2000

C:Accession: B38345; JU0123

R:Shigenaga, T.; Muta, T.; Toh, Y.; Tokunaga, F.; Iwanaga, S.

J. Biol. Chem. 265, 21350-21354, 1990

A:Title: Antimicrobial tachyplesin peptide precursor. cDNA cloning and cellular localization

A:Reference number: A38345; MUID:91065956

A:Accession: B38345

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-77 <SHI>

C:Cross-references: GB:J05689

R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J.

Biochem. 106, 663-668, 1989

A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin

A:Reference number: A91914; MUID:90110066

A:Accession: JU0123

A:Molecule type: protein

A:Residues: 24-40 <MIY>

C:Comment: The peptide is one of the antimicrobial peptides found in the Japanese horseshoe crab

C:Keywords: amidated carboxyl end

F:26-39,30-35/Disulfide bonds: #status predicted

F:40/Modified site: amidated carboxyl end (Arg) (amide in mature form from following

Query Match 71.4%; Score 65; DB 2; Length 77;

Best Local Similarity 64.7%; Pred. No. 0.0007;

Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||| ||| ||| |||

Db 24 KWCFRVCYRGICYRRCR 40

# RESULT 6

JU0124

polyphemus I - Atlantic horseshoe crab

C:Species: Limulus polyphemus (Atlantic horseshoe crab)

C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jul-1997

C:Accession: JU0124

R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J.

Biochem. 106, 663-668, 1989

A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin

A:Reference number: A91914; MUID:90110066

A:Accession: JU0124

A:Molecule type: protein

A:Residues: 1-18 <MIY>

C:Comment: The peptide is one of the antimicrobial peptides in the Atlantic horseshoe crab

C:Keywords: amidated carboxyl end

F:4-17,8-13/Disulfide bonds: #status experimental

F:18/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 67.0%; Score 61; DB 2; Length 18;

Best Local Similarity 58.8%; Pred. No. 0.00074;

Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||| ||| ||| |||

Db 2 KWCFRVCYRGICYRRCR 18

# RESULT 7

JU0125

polyphemus II - Atlantic horseshoe crab

C:Species: Limulus polyphemus (Atlantic horseshoe crab)

C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jul-1997

C:Accession: JU0125

R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J.

Biochem. 106, 663-668, 1989

A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin

A:Reference number: A91914; MUID:90110066

A:Accession: JU0125

A:Molecule type: protein

A:Residues: 1-18 <MIY>  
 C:Comment: The peptide is one of the antimicrobial peptides in the American horseshoe crab  
 C:Keywords: amidated carboxyl end  
 F:4-17, 8-13/Disulfide bonds: #status predicted  
 F:18/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 63.7%; Score 58; DB 2; Length 18;  
 Best Local Similarity 52.9%; Pred. No. 0.0023;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
 :| ||| | :| :|  
 Db 2 RWCFRVCYKGFYKCR 18

RESULT 8  
 D86473  
 hypothetical protein AAG27097.1 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: D86473  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: D86473  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-615 <STO>  
 A:Cross-references: GB:AE005172; NID:gl1034940; PIDN:AAG27097.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 51.6%; Score 47; DB 2; Length 615;  
 Best Local Similarity 56.2%; Pred. No. 5.3;  
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRS 16  
 :| ||| ||| :| :|  
 Db 175 QWFRHSYRGTPQRHS 190

RESULT 9  
 C82472  
 conserved hypothetical protein VCA0332 [imported] - Vibrio cholerae (strain N16961 serog  
 C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: C82472  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H  
 I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 405, 477-483, 2000  
 A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833  
 A:Accession: C82472  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-88 <HEI>  
 A:Cross-references: GB:AE004372; GB:AE003853; NID:g9657741; PIDN:AAF96240.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VCA0332  
 A:Map position: 2

Query Match 48.9%; Score 44.5; DB 2; Length 88;  
 Best Local Similarity 52.4%; Pred. No. 1.9;  
 Matches 11; Conservative 2; Mismatches 3; Indels 5; Gaps 1;

QY 2 WSRFVSYRG-----ISYRRSR 17  
 || :||| || ||||  
 Db 58 WSGVITYRGTNIRIISVRRSR 78

RESULT 10  
 F84162  
 hypothetical protein Vng0026c [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: F84162  
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Las  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.;  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483  
 A:Accession: F84162  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-307 <STO>  
 A:Cross-references: GB:AE004437; NID:gl0579674; PIDN:AAG186656.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG0026C

Query Match 48.4%; Score 44; DB 2; Length 307;  
 Best Local Similarity 42.9%; Pred. No. 8.2;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYR 14  
 ||:||| | :| :|  
 Db 181 KNAFRTLYEQVAYK 194

RESULT 11  
 A84161  
 hypothetical protein Vng0013c [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: A84161  
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Las  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.;  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483  
 A:Accession: A84161  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-378 <STO>  
 A:Cross-references: GB:AE004437; NID:gl0579658; PIDN:AAG18653.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG0013C

Query Match 48.4%; Score 44; DB 2; Length 378;  
 Best Local Similarity 42.9%; Pred. No. 10;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYR 14  
 | :||| | :| :|  
 Db 252 KNAFRTLYEQVAYK 265

RESULT 12

E86468  
protein F12K21.26 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E86468  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E86468  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-620 <STO>  
A:Cross-references: GB:AE005172; NID:g8778254; PIDN:AAF79263.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F12K21.26  
A:Map position: 1

Query Match 48.4%; Score 44; DB 2; Length 620;  
Best Local Similarity 50.0%; Pred. No. 17;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
QY 1 KWSFRVSYRGISYRRS 16  
|||:|:|:|:|:|  
DB 150 QMRFRHNYRGTPQRHS 165  
|||:|:|:|:|:|  
RESULT 13  
molybdate metabolism regulator-related protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: G75313  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: G75313  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-480 <WHI>  
A:Cross-references: GB:AE002046; GB:AE000513; NID:g6459901; PIDN:AAF11657.1; PID:g645990  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2108  
A:Map position: 1

Query Match 46.2%; Score 42; DB 2; Length 480;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
QY 2 WSRFRVSYRGISYRRSR 17  
|:|:|:|:|:|:|  
DB 234 WMVRVDARGVYGHSR 249  
|:|:|:|:|:|:|  
RESULT 14  
D42463  
hypothetical protein Bcv' (pinB 5' region) - Shigella boydii (fragment)  
C:Species: Shigella boydii  
C:Date: 10-Jul-1992 #sequence\_revision 18-Sep-1992 #text\_change 30-Sep-1993  
-C:Accession: D42463

R:Tominaga, A.; Ikemizu, S.; Enomoto, M.  
J. Bacteriol. 173, 4079-4087, 1991  
A:Title: Site-specific recombinase genes in three Shigella subgroups and nucleotide  
A:Reference number: A42463; MUID:91286192  
A:Accession: D42463  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-536 <TOM>  
A:Cross-references: GB:D00660

Query Match 46.2%; Score 42; DB 2; Length 536;  
Best Local Similarity 62.5%; Pred. No. 31;  
Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;  
QY 4 FRVSYR--GISYRRSR 17  
|||:|:|:|:|:|:|  
DB 423 FRVNYRNGGIFYRSAR 438  
|||:|:|:|:|:|:|  
RESULT 15  
T02495  
hypothetical protein At2g38500 [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein T19C21.1  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001  
C:Accession: T02495; G84805  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.;  
submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.  
A:Reference number: Z14676  
A:Accession: T02495  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-279 <ROU>  
A:Cross-references: EMBL:AC004683; NID:g3395421; PID:g3395422  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vente  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: G84805  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-279 <STO>  
A:Cross-references: GB:AE002093; NID:g3786022; PIDN:AAC67368.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: T19C21.1; At2g38500  
A:Map position: 2  
A:Introns: 170/2

Query Match 45.1%; Score 41; DB 2; Length 279;  
Best Local Similarity 80.0%; Pred. No. 23;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 SFRVSRYRGIS 12  
:|:|:|:|:|:|  
DB 99 AFRVSYHGIS 108  
:|:|:|:|:|:|

Search completed: February 12, 2002, 12:34:41  
Job time: 559 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:52 : Search time 67.2 Seconds  
(without alignments)  
9.275 Million cell updates/sec

Title: US-09-485-571-26

Perfect score: 91

Sequence: 1 KWSFRVSYRGISYRRSR 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	71	78.0	17	TAC1_TACGI	P23684 tachypleus
2	71	78.0	77	TAC1_TACTR	P14213 tachypleus
3	68	74.7	17	TAC3_TACGI	P18252 tachypleus
4	65	71.4	77	TAC2_TACTR	P14214 tachypleus
5	61	67.0	18	PPM1_LIMPO	P14215 limulus pol
6	58	63.7	18	PPM2_LIMPO	P14216 limulus pol
7	40	44.0	473	DNB2_ADE40	P11806 human adeno
8	40	44.0	474	DNB2_ADE41	P11807 human adeno
9	40	44.0	1100	JAK3_RAT	Q63272 rattus norv
10	40	44.0	1124	JAK3_HUMAN	P52333 homo sapien
11	40	44.0	1299	JAK3_MOUSE	O62117 mus musculu
12	39	42.9	271	URED_ACTNA	Q92364 actinomycet
13	39	42.9	968	BCAL_RAT	P42512 pseudomonas
14	38	41.8	203	L2CC_DROME	P33238 anas platyr
15	38	41.8	261	HB22_HUMAN	Q00662 dianthus ca
16	38	41.8	411	LAG1_YEAST	RL11_ORYSA
17	38	41.8	484	DNB2_ADE12	O22540 oryza sativ
18	38	41.8	616	ENP4_HUMAN	P09237 homo sapien
19	38	41.8	720	FPTA_PSEAE	P20282 bacillus su
20	38	41.8	721	MX_ANAPL	P49006 homo sapien
21	38	41.8	731	BGAL_DIAEA	P35566 oryctolagus
22	37.5	41.2	182	RL11_ORYSA	P28667 mus musculu
23	37.5	41.2	267	MM07_HUMAN	P36113 saccharomyc
24	37	40.7	120	RS13_BACSU	P22956 homo sapien
25	37	40.7	194	MRP_HUMAN	P22967 mus musculu
26	37	40.7	198	MRP_RABIT	P38456 marchantia
27	37	40.7	199	MRP_MOUSE	P22968 oryctolagus
28	37	40.7	551	YK27_YEAST	Q08047 zea mays (m
29	37	40.7	732	ACET_HUMAN	
30	37	40.7	732	ACET_MOUSE	
31	37	40.7	732	YMI1_MARPO	
32	37	40.7	737	ACET_RABIT	
33	37	40.7	799	GLGB_MAIZE	

34 37 40.7 1306 1 ACE\_HUMAN  
35 37 40.7 1310 1 ACE\_RABIT  
36 37 40.7 1312 1 ACE\_MOUSE  
37 37 40.7 1313 1 ACE\_RAT  
38 37 40.7 1453 1 NKCR\_MOUSE  
39 37 40.7 1462 1 NKCR\_HUMAN  
40 36.5 40.1 181 1 RL11\_MEDSA  
41 36.5 40.1 182 1 RL11\_ARATH  
42 36.5 40.1 182 1 RL11\_ARATH  
43 36.5 40.1 184 1 RL12\_ARATH  
44 36.5 40.1 671 1 AMOL\_ASPNG  
45 36 39.6 90 1 IATP\_SCHPO

P12821 homo sapien  
P12822 oryctolagus  
P09470 mus musculu  
P47820 rattus norv  
P30415 mus musculu  
P30414 homo sapien  
P46287 medicago sa  
P42795 arabidopsis  
P42794 arabidopsis  
P42796 arabidopsis  
Q12556 aspergillus  
O74523 schizosacch

#### ALIGNMENTS

RESULT 1  
TAC1\_TACGI  
ID TAC1\_TACGI STANDARD; PRT; 17 AA.  
AC P23684;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE TACHYPLESIN I.  
OS Tachypleus gigas (Southeast Asian horseshoe crab), and  
OC Cariniscorpius rotundicauda (Southeast Asian horseshoe crab).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
OC Limulidae; Tachypleus.  
OX NCBI\_TaxID=6852, 6848;  
RN [1]  
RP SEQUENCE.  
RC SPECIES-T. gigas, and C. rotundicauda;  
RX MEDLINE=91035357; Pubmed=2229025;  
RA Muta T., Fujimoto T., Nakajima H., Iwanaga S.;  
RT "Tachypleus isolated from hemocytes of Southeast Asian horseshoe  
crabs (Cariniscorpius rotundicauda and Tachypleus gigas);  
RT identification of a new tachypleusin, tachypleusin III, and a  
RT processing intermediate of its precursor";  
RL J. Biochem. 108:261-266(1990).  
CC -!- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
DR PIR; A38824; A38824.  
DR PIR; JX0124; JX0124.  
KW Antibiotic; Amidation.  
FT DISULFID 3 16  
FT DISULFID 7 12  
FT MOD\_RES 17 17  
SQ SEQUENCE 17 AA; 2269 MW; E9E09BD9D2923C94 CRC64;  
Query Match 78.0% Score 71; DB 1; Length 17;  
Best Local Similarity 76.5% Pred. No. 3.4e-06;  
Matches 13; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 KWSFRVSYRGISYRRSR 17  
II III IIII IIII  
Db 1 KWCFCVRCYRCYRRCR 17  
RESULT 2  
TAC1\_TACTR  
ID TAC1\_TACTR STANDARD; PRT; 77 AA.  
AC P14213;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE TACHYPLESIN I PRECURSOR.  
OS Tachypleus tridentatus (Japanese horseshoe crab).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
OC Limulidae; Tachypleus.  
OX NCBI\_TaxID=6853;  
RN [1]  
RP SEQUENCE FROM N.A.

Matches 13; Conservative 0; Mismatches 4; Indels 0; Caps 0

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||| ||| ||| |

Db 24 KWCFRVCYRGICYYRKR 40

RESULT 3

TAC3\_TACGI  
 ID TAC3\_TACGI STANDARD; PRT; 17 AA.

AC P18252;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE TACHYPLESIN III.  
 DE TACHYPLESIN gigas (Southeast Asian horseshoe crab).  
 OS Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
 OC Limulidae; Tachyplesus.  
 OC NCBI\_TaxID=6852;  
 OX  
 RN [1]  
 RP SEQUENCE.  
 RP MEDLINE=91035357; PubMed=2229025;  
 RX Muta T., Fujimoto T., Nakajima H., Iwanaga S.;  
 RA "Tachyplesins isolated from hemocytes of Southeast Asian horseshoe  
 crabs (Carcinoscopus rotundicauda and Tachyplesus gigas):  
 RT identification of a new tachyplesin, tachyplesin III, and a  
 RT processing intermediate of its precursor.";  
 RT J. Biochem. 108:261-266(1990).  
 CC -1- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND  
 CC -1- GRAM-POSITIVE BACTERIA.  
 CC -1- TISSUE SPECIFICITY: HEMOCYTES.  
 CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
 CC PIR: JX0125; JX0125.  
 KW Antibiotic; Amidation.  
 FT DISULFID 3 16 BY SIMILARITY.  
 FT DISULFID 7 12 BY SIMILARITY.  
 FT MOD\_RES 17 17 AMIDATION  
 FT SEQUENCE 17 AA; 2241 MW; E9E08CE9D2923C94 CRC64;

Query Match 74.7%; Score 68; DB 1; Length 17;  
 Best Local Similarity 70.6%; Pred. No. 1.1e-05;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Caps 0

Qy 1 KWSFRVSYRGISYRRSR 17  
 || ||| ||| ||| |

Db 1 KWCFRVCYRGICYYRKR 17

RESULT 4

TAC2\_TACTR  
 ID TAC2\_TACTR STANDARD; PRT; 77 AA.

AC P14214;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE TACHYPLESIN II PRECURSOR.  
 DE Tachyplesus tridentatus (Japanese horseshoe crab).  
 OS Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
 OC Limulidae; Tachyplesus.  
 OC NCBI\_TaxID=6853;  
 OX  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=91065956; PubMed=2250028;  
 RX Shigenaga T., Muta T., Toh Y., Tokunaga F., Iwanaga S.;  
 RA "Antimicrobial tachyplesin peptide precursor. cDNA cloning and  
 RT cellular localization in the horseshoe crab (Tachyplesus  
 RT tridentatus).";  
 RT J. Biol. Chem. 265:21350-21354(1990).  
 RN [2]  
 RP SEQUENCE OF 24-40.  
 RX MEDLINE=90110066; PubMed=2514185;  
 RA Miyata T., Tokunaga F., Yonega T., Yoshikawa K., Iwanaga S., Niwa M.,



CC INDEPENDENT MANNER. INVOLVED IN DNA-REPLICATION, REGULATION OF  
 CC MRNA FORMATION, AND HOST-RANGE SPECIFICITY. ZINC IS REQUIRED  
 CC FOR DNA BINDING.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. ACCUMULATES IN INFECTED CELLS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M19540; AAA52196.1; -  
 CC EMBL: L19443; AAC13969.1; -  
 CC EMBL: A28645; ERAD40.  
 CC HSP: P03265; IADV.  
 CC InterPro: IPR003176; Vir\_DNA\_binding.  
 CC Pfam: PF02236; Vir\_DNA\_binding; 1.  
 CC Early protein; DNA-binding; Zinc-finger; Phosphorylation;  
 CC Nuclear protein. 141 PHOSPHORYLATION (PROBABLE).  
 CC MOD\_RES 141 141  
 CC ZN\_FING 219 232 POTENTIAL.  
 CC SEQUENCE 473 AA; 5335 MW; 8273635BF6703A24 CRC64;  
 CC  
 CC Query Match 44.0%; Score 40; DB 1; Length 473;  
 CC Best Local Similarity 50.0%; Pred. No. 20;  
 CC Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 CC  
 CC QY 1 KWSFRVSYRGIS 12  
 CC III I: II :  
 CC Db 449 KWSRLQYRNVA 460  
 CC  
 CC RESULT 8  
 CC DNB2\_ADE41  
 CC ID DNB2\_ADE41 STANDARD; PRT; 474 AA.  
 CC AC P11807;  
 CC DT 01-OCT-1989 (Rel. 12, Created)  
 CC DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 CC DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 CC DE EARLY E2A DNA-BINDING PROTEIN.  
 CC GN DBP.  
 CC OS Human adenovirus type 41.  
 CC OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 CC OX NCBI\_TaxID=10524;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=88160034; PubMed=3279700;  
 CC Vos H.L., der Lee F.M., Reemst A.M.C.B., van Loon A.E.,  
 CC Sussenbach J.S.;  
 CC "The genes encoding the DNA binding protein and the 23K protease of  
 CC adenovirus types 40 and 41";  
 CC Virology 163:1-10(1986).  
 CC [2]  
 CC SEQUENCE OF 1-33 FROM N.A.  
 CC STRAIN=TAK;  
 CC MEDLINE=90272433; PubMed=2349115;  
 CC Slemenda S.B., Pieniazek N.J., Velarde J. Jr., Pieniazek D.,  
 CC Luftig R.B.;  
 CC "Nucleotide sequence of the region coding for 100K and 33K proteins  
 CC of human enteric adenovirus type 41 (Tak).";  
 CC Nucleic Acids Res. 18:3069-3069(1990).  
 CC -!- FUNCTION: BINDS COOPERATIVELY SINGLE-STRANDED DNA IN A SEQUENCE-  
 CC INDEPENDENT MANNER. INVOLVED IN DNA-REPLICATION, REGULATION OF  
 CC MRNA FORMATION, AND HOST-RANGE SPECIFICITY. ZINC IS REQUIRED  
 CC FOR DNA BINDING.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. ACCUMULATES IN INFECTED CELLS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M21163; AAA2463.1; -  
 CC EMBL: X52532; CAA36759.1; -  
 CC PIR: D28645; ERAD41.  
 CC PIR: S10206; S10206.  
 CC HSP: P03265; IADV.  
 CC InterPro: IPR003176; Vir\_DNA\_binding.  
 CC Pfam: PF02236; Vir\_DNA\_binding; 1.  
 CC Early protein; DNA-binding; Zinc-finger; Phosphorylation.  
 CC MOD\_RES 142 142 PHOSPHORYLATION (PROBABLE).  
 CC ZN\_FING 220 233 POTENTIAL.  
 CC SEQUENCE 474 AA; 53658 MW; 4350AES93088B19E CRC64;  
 CC  
 CC Query Match 44.0%; Score 40; DB 1; Length 474;  
 CC Best Local Similarity 50.0%; Pred. No. 20;  
 CC Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 CC  
 CC QY 1 KWSFRVSYRGIS 12  
 CC III I: II :  
 CC Db 450 KWSRLQYRNVA 461  
 CC  
 CC RESULT 9  
 CC JAK3\_RAT  
 CC ID JAK3\_RAT STANDARD; PRT; 1100 AA.  
 CC AC O63272;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 CC DE TYROSINE-PROTEIN KINASE JAK3 (EC 2.7.1.112) (JANUS KINASE 3) (JAK-3).  
 CC GN JAK3.  
 CC OS Rattus norvegicus (Rat).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC OX NCBI\_TaxID=10116;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC TISSUE=Spleen;  
 CC MEDLINE=94192816; PubMed=8143863;  
 CC Takahashi T., Shirasawa T.;  
 CC "Molecular cloning of rat JAK3, a novel member of the JAK family of  
 CC protein tyrosine kinases";  
 CC FEBS Lett. 342:124-128(1994).  
 CC -!- FUNCTION: TYROSINE KINASE OF THE NON-RECEPTOR TYPE, INVOLVED IN  
 CC THE INTERLEUKIN-2 AND INTERLEUKIN-4 SIGNALING PATHWAY.  
 CC PHOSPHORYLATES STAT6, IRS1, IRS2 AND PI3K.  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE -> ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: WHOLLY INTRACELLULAR, POSSIBLY MEMBRANE  
 CC ASSOCIATED (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: TRANSCRIBED IN A VARIETY OF TISSUES INCLUDING  
 CC SPLEEN, LUNG, KIDNEY AND INTESTINE.  
 CC -!- DOMAIN: POSSESSES TWO PHOSPHOTRANSFERASE DOMAINS. THE SECOND ONE  
 CC PROBABLY CONTAINS THE CATALYTIC DOMAIN (BY SIMILARITY), WHILE THE  
 CC PRESENCE OF SLIGHT DIFFERENCES SUGGEST A DIFFERENT ROLE FOR  
 CC DOMAIN 1.  
 CC -!- PTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IL-2 AND IL-4 (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. JAK  
 CC SUBFAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: D28508; BAA05868.1; -.
DR HSSP: P11362; IFCI.
DR InterPro: IPR000299; Band_4.1.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001245; Tyr_Kin.
DR Pfam: PF00069; pkinase; 2.
DR SMART: SM00295; B41.1.
DR SMART: SM00252; SH2.1.
DR SMART: SM00219; TykKc.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 2.
DR PROSITE: PS00001; SH2_FALSE_NEG.
DR Transfaser: Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW SH2 domain; Repeat.
FT DOMAIN 372 472 SH2 (ATYPICAL).
FT DOMAIN 517 777 PROTEIN_KINASE 1.
FT DOMAIN 818 1091 PROTEIN_KINASE 2.
FT NP_BIND 824 832 ATP (BY SIMILARITY).
FT BINDING 851 851 ATP (BY SIMILARITY).
FT ACT_SITE 945 945 BY SIMILARITY..
FT MOD_RES 976 976 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 1100 AA; 122560 MW; 1D59CA05F4DD7EE2 CRC64;

Query Match 44.08; Score 40; DB 1; Length 1100;
Best Local Similarity 57.19; Pred. No. 47;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 FRVSYRGISYRRSR 17
D 878 FIVXRGVSYGGR 891
I I I I I I I I I
I I I I I I I I I

RESULT 10
JAK3_HUMAN
ID JAK3_HUMAN STANDARD; PRT; 1124 AA.
AC P52333; Q13259; Q13260; Q13611;
DT 01-OCT-1996 (Rel. 34, Created);
DT 01-OCT-1996 (Rel. 34, Last sequence update);
DT 20-AUG-2001 (Rel. 40, Last annotation update);
DE TYROSINE-PROTEIN KINASE JAK3 (EC 2.7.1.112) (JANUS KINASE 3) (JAK-3)
DE (LEUKOCYTE JANUS KINASE) (L-JAK).
GN JAK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94294384; PubMed=8022790;
RA Kawamura M., McVicar D.W., Johnston J.A., Blake T.B., Chen Y.-Q.,
RA Lal B.K., Lloyd A.R., Kelvin D.J., Staples J.E., Ortaldo J.R.,
RA O'Shea J.J.;
RT "Molecular cloning of L-JAK, a Janus family protein-tyrosine kinase
RT expressed in natural killer cells and activated leukocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:6374-6378(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96027605; PubMed=7559633;
RA Lal K.S., Jin Y., Graham D.K., Witthuhn B.A., Ihle J.N., Liu E.T.;
RT "A kinase-deficient splice variant of the human JAK3 is expressed in
RT hematopoietic and epithelial cancer cells.";
RL J. Biol. Chem. 270:25028-25036(1995).
RN [3]
RP SEQUENCE OF 36-191 FROM N.A.
RX MEDLINE=96278845; PubMed=8662778;
RA Verbsky J.W., Bach E.A., Fang L., Randolph D.A.,
RA Fields L.E.;
RT "Expression of Janus kinase 3 in human endothelial and other non-
```

```

RT lymphoid and non-myeloid cells.";
RL J. Biol. Chem. 271:13976-13980(1996).
RN [4]
RP VARIANT SCID CVS-100.
RX MEDLINE=95388142; PubMed=7659163;
RA Macchi P., Villa A., Gilliani S., Sacco M.G., Frattini A., Porta F.,
RA Ugazio A.G., Johnston J.A., Candotti F., O'Shea J.J., Vezzoni P.,
RA Notarangelo L.D.;
RT "Mutations of Jak-3 gene in patients with autosomal severe combined
RT immune deficiency (SCID).";
RL Nature 377:65-68(1995).
RN [5]
RP VARIANTS SCID GLY-481; LEU-586--MET-592 DEL AND ARG-759.
RX MEDLINE=98022793; PubMed=9354668;
RA Candotti F., Oakes S.A., Johnston J.A., Gilliani S., Schumacher R.F.,
RA Mella P., Florini M., Ugazio A.G., Badolato R., Notarangelo L.D.,
RA Bozzi F., Macchi P., Strina D., Vezzoni P., Blaese R.M., O'Shea J.J.,
RA Villa A.;
RT "Structural and functional basis for JAK3-deficient severe combined
RT immunodeficiency.";
RL Blood 90:3996-4003(1997).
RN [6]
RP VARIANT SCID TRP-582.
RX MEDLINE=98433994; PubMed=9753072;
RA Bozzi F., Lefranc G., Villa A., Badolato R., Schumacher R.F.,
RA Khalil G., Loiselet J., Bresciani S., O'Shea J.J., Vezzoni P.,
RA Notarangelo L.D., Candotti F.;
RT "Molecular and biochemical characterization of JAK3 deficiency in a
RT patient with severe combined immunodeficiency over 20 years after
RT bone marrow transplantation: implications for treatment.";
RL Br. J. Haematol. 102:1363-1366(1998).
RN [7]
RP VARIANTS SCID ARG-151; ILE-722 AND SER-910.
RX MEDLINE=20435064; PubMed=10982185;
RA Schumacher R.F., Mella P., Badolato R., Florini M., Savoldi G.,
RA Gilliani S., Villa A., Candotti F., Tampalini A., O'Shea J.J.,
RA Notarangelo L.D.;
RT "Complete genomic organization of the human JAK3 gene and mutation
RT analysis in severe combined immunodeficiency by single-strand
RT conformation polymorphism.";
RL Hum. Genet. 106:73-79(2000).
CC -1- FUNCTION: TYROSINE KINASE OF THE NON-RECEPTOR TYPE, INVOLVED IN
CC THE INTERLEUKIN-2 AND INTERLEUKIN-4 SIGNALING PATHWAY.
CC PHOSPHORYLATES STAT6, IRS1, IRS2 AND PI3K.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: WHOLLY INTRACELLULAR, POSSIBLY MEMBRANE
CC ASSOCIATED (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: THREE SPLICE VARIANTS WERE ISOLATED FROM
CC DIFFERENT MRNA SOURCES: BREAST (JAK3B), SPLEEN (JAK3S; SHOWN
CC HERE), AND ACTIVATED MONOCYTES (JAK3M). JAK3B MAY BE DEFFECTIVE
CC AS IT LACK SOME PART OF THE KINASE DOMAIN.
CC -1- TISSUE SPECIFICITY: IN NK CELLS AND AN NK-LIKE CELL LINE BUT NOT
CC IN RESTING T CELLS OR IN OTHER TISSUES. THE S-FORM IS MORE
CC COMMONLY SEEN IN HEMATOPOIETIC LINES, WHEREAS THE B- AND M-FORMS
CC ARE DETECTED IN CELLS BOTH OF HEMATOPOIETIC AND EPITHELIAL
CC ORIGINS.
CC -1- DOMAIN: POSSESSES TWO PHOSPHOTRANSFERASE DOMAINS. THE SECOND ONE
CC PROBABLY CONTAINS THE CATALYTIC DOMAIN (BY SIMILARITY), WHILE THE
CC PRESENCE OF SLIGHT DIFFERENCES SUGGEST A DIFFERENT ROLE FOR
CC DOMAIN 1.
CC -1- PTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IL-2 AND IL-4.
CC -1- DISEASE: DEFECTS IN JAK3 ARE A CAUSE OF RECESSIVE T-CELL
CC NEGATIVE/B-CELL POSITIVE SEVERE COMBINED IMMUNODEFICIENCY (T-B+
CC SCID), A CONDITION CHARACTERIZED BY THE ABSENCE OF CIRCULATING
CC MATURE T-LYMPHOCYTES AND NK CELLS, NORMAL TO ELEVATED NUMBERS OF
CC NONFUNCTIONAL B-LYMPHOCYTES, AND MARKED HYPOPLASIA OF LYMPHOID
CC TISSUES.
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. JAK
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; U09607; AAC19626.1; -  
DR EMBL; U31601; AAC50226.1; -  
DR EMBL; U31602; AAC50227.1; -  
DR EMBL; U57096; AAC50342.1; -  
DR HSSP; P11382; 1FGI.  
DR MIM; 600173; -  
DR MIM; 600802; -  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR000980; SH2.  
DR InterPro; IPR001245; Tyr\_kin.  
DR Pfam; PF00069; pkinase; 2.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00252; SH2; 1.  
DR SMART; SM00219; TYRKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 2.  
DR PROSITE; PS00001; SH2; FALSE\_NEG.  
DR Transferase; Tyrosine-protein kinase: ATP-binding; Phosphorylation;  
KW SH2 domain; Repeat; Alternative splicing; Disease mutation; SCID.  
KW DOMAIN 375 475 SH2 (ATYPICAL).  
FT DOMAIN 521 781 PROTEIN KINASE 1.  
FT DOMAIN 822 1111 PROTEIN KINASE 2.  
FT NP\_BIND 828 836 ATP (BY SIMILARITY).  
FT BINDING 855 855 ATP (BY SIMILARITY).  
FT BINDING 855 855 BY SIMILARITY.  
FT ACT\_SITE 949 949 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 980 980 HELMKLCWAPSQDPRPSFSAALGQPLDMLWSGRGCEHAF.  
FT VARSPPLIC 1071 1124 AHPEGKHLSLFS -> SAAGLASVSQSDVMAGVSGKPAGA  
FT FT (IN ISOFORM JAK3B).  
FT FT HELMKLCWAPSQDPRPSFSAALGQPLDMLWSGRGCEHAF  
FT FT AHPEGKHLSLFS -> SCYSGWRDDICSMGWPTVIRWD  
FT FT LACSPCPRLTATVTLQPLPPLTHATAASVAVPNKTC (IN  
FT FT ISOFORM JAK3M).  
FT FT Y -> C (IN SCID).  
FT FT /FTid=VAR\_006284.  
FT FT P -> R (IN SCID).  
FT FT /FTid=VAR\_010492.  
FT FT E -> G (IN SCID).  
FT FT /FTid=VAR\_010493.  
FT FT R -> W (IN SCID).  
FT FT /FTid=VAR\_010494.  
FT FT MISSING (IN SCID); LACK OF PHOSPHORYLATION  
FT FT IN RESPONSE TO CYTOKINE STIMULATION).  
FT FT /FTid=VAR\_010495.  
FT FT V -> I (IN SCID).  
FT FT /FTid=VAR\_010496.  
FT FT C -> R (IN SCID); CONSTITUTIVE  
FT FT PHOSPHORYLATION).  
FT FT /FTid=VAR\_010497.  
FT FT L -> S (IN SCID).  
FT FT /FTid=VAR\_010498.  
FT FT A -> G (IN REF. 2).  
FT FT MISSING (IN REF. 3).  
FT FT T -> A (IN REF. 3).  
FT FT A -> R (IN REF. 2).  
FT FT P -> R (IN REF. 2).  
FT FT M -> I (IN REF. 2).  
FT FT AH -> GD (IN REF. 2).  
FT FT PE -> QS (IN REF. 2).  
FT FT 1124 AA; 125015 MW; 1D0FD22068E088E4 CRC64;  
SQ SEQUENCE

Query Match 44.0%; Score 40; DB 1; Length 1124;  
Best Local Similarity 57.1%; Pred. No. 48;

QY 4 FRVYRGISYRRSR 17  
| | | | | | | | | |  
Db 882 FIVKYGVSYGPCR 895

RESULT 11  
JAK3\_MOUSE  
ID JAK3\_MOUSE STANDARD; PRT: 1299 AA.  
AC Q62137; Q61747; Q61746;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TYROSINE-PROTEIN KINASE JAK3 (EC 2.7.1.112) (JANUS KINASE 3) (JAK-3).  
GN JAK3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RP MEDLINE=94309920; PubMed=7518579;  
RX Rane S.G., Reddy E.P.;  
RA "JAK3: a novel JAK kinase associated with terminal differentiation of  
RT hematopoietic cells.";  
RL Oncogene 9:2415-2423(1994).  
[2]  
RN SEQUENCE FROM N.A. (ISOFORM 2).  
RP STRAIN=BALB/C X 129 F2; TISSUE=Thymus;  
RX MEDLINE=96184772; PubMed=8605329;  
RA Gurniak C.B., Berg L.J.;  
RT "Murine JAK3 is preferentially expressed in hematopoietic tissues and  
RT lymphocyte precursor cells.";  
RL Blood 87:3151-3160(1996).  
[3]  
RN SEQUENCE FROM N.A. (ISOFORM 3).  
RP STRAIN=BALB/C;  
RX MEDLINE=94294024; PubMed=8022486;  
RA Withuhn B.A., Silvennoinen O., Mlura O., Lai K.S., Cwik C., Liu E.T.,  
RA Ihle J.N.;  
RT "Involvement of the Jak-3 Janus kinase in signalling by interleukins  
RT 2 and 4 in lymphoid and myeloid cells.";  
RL Nature 370:153-157(1994).  
CC -!- FUNCTION: TYROSINE KINASE OF THE NON-RECEPTOR TYPE, INVOLVED IN  
CC THE INTERLEUKIN-2 AND INTERLEUKIN-4 SIGNALING PATHWAY.  
CC PHOSPHORYLATES STAT6, IRS1, IRS2 AND PI3K.  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: WHOLLY INTRACELLULAR, POSSIBLY MEMBRANE  
CC ASSOCIATED (BY SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN THE THYMUS WITH  
CC SOMEWHAT LOWER LEVELS IN BONE MARROW, SPLEEN, FETAL LIVER AND  
CC ADULT CD4-CD8- THYMOCYTES. VERY LOW LEVELS IN ADULT KIDNEY, LUNG,  
CC TESTES, BRAIN AND LIVER.  
CC -!- DOMAIN: POSSESSES TWO PHOSPHOTRANSFERASE DOMAINS. THE SECOND ONE  
CC PROBABLY CONTAINS THE CATALYTIC DOMAIN (BY SIMILARITY), WHILE THE  
CC PRESENCE OF SLIGHT DIFFERENCES SUGGEST A DIFFERENT ROLE FOR  
CC DOMAIN 1.  
CC -!- PTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IL-2 AND IL-4.  
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. JAK  
CC SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
-----
CC EMBL; L33768; AAA21415.1; -
DR EMBL; L40172; AAC42085.1; -
DR EMBL; L32955; AAA21565.1; -
DR HSSP; P11362; IFGI
DR MGD; MGI:99928; Jak3
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001245; Tyr_Kin.
DR Pfam; PF00069; pkinase; 3.
DR SMART; SM00295; B41; 1.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00219; Tyfkc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS50001; SH2; FALSE_NEG.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW SH2 domain; Repeat; Alternative splicing.
FT DOMAIN 560 688 SH2 (ATYPICAL).
FT DOMAIN 732 992 PROTEIN KINASE 1.
FT DOMAIN 1032 1299 PROTEIN KINASE 2.
FT NP_BIND 1038 1046 ATP (BY SIMILARITY).
FT BINDING 1064 1064 ATP (BY SIMILARITY).
FT ACT_SITE 1158 1158 BY SIMILARITY.
FT MOD_RES 1189 1189 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT VARSPPLIC 222 238 RVMSPARPTATAHGQVY -> LPGRLPGRPYALMKYI
FT VARSPPLIC 223 238 VMSPARPTATAHGQVY -> RVACQADRYI (IN
FT VARSPPLIC 284 473 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPPLIC 491 498 OPTCGSGR -> QAPRVGPAG (IN ISOFORM 2 AND
FT VARSPPLIC 568 596 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPPLIC 656 673 ASASPTACGSCQLLEF -> GLSQPHRSRLRELLAACWNS
FT VARSPPLIC 1000 1027 QTPHLASRVLEMSCAWRPALCLPGPRHI -> SDPTPGIPS
FT VARSPPLIC 1290 1299 SRTGQPSAP -> EPHDRPAFATLSQOLDPLWRGPG
FT MUTAGEN 1066 1066 L->R: LOSS OF ACTIVITY.
FT CONFLICT 62 62 A -> G (IN REF. 2).
FT CONFLICT 276 276 S -> P (IN REF. 3).
FT CONFLICT 280 280 G -> N (IN REF. 3).
FT CONFLICT 282 282 MISSING (IN REF. 3).
FT CONFLICT 490 490 K -> N (IN REF. 3).
FT CONFLICT 550 550 A -> P (IN REF. 2 AND 3).
FT CONFLICT 559 560 EL -> DV (IN REF. 3).
FT CONFLICT 607 607 A -> G (IN REF. 2 AND 3).
FT CONFLICT 683 683 N -> Y (IN REF. 3).
FT CONFLICT 706 706 T -> N (IN REF. 3).
FT CONFLICT 734 734 G -> EW (IN REF. 2 AND 3).
FT CONFLICT 750 750 R -> S (IN REF. 3).
FT CONFLICT 931 932 SG -> OR (IN REF. 3).
SQ SEQUENCE 1299 AA; 144314 MW; EFE2D60B6AF3D10C CRC64;
```

Query Match 44.0%; Score 40; DB 1; Length 1299;  
Best Local Similarity 57.1%; Pred. No. 56;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 FRVSRGSIYRRSR 17  
| | | | | | | |  
DB 1091 FIVKRGVSGPGR 1104

RESULT 12  
URED\_ACTNA STANDARD; PRT; 271 AA.  
AC Q92364;  
DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE UREASE ACCESSORY PROTEIN URED.  
GN URED.  
OS Actinomyces naeslundii.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.  
OX NCBI\_TaxID=1655;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VVU45;  
RX MEDLINE=99115518; PubMed=9916052;  
RA Morou-Bermudez E., Burne R.A.;  
RT "Genetic and physiologic characterization of urease of Actinomyces  
naeslundii.";  
RL Infect. Immun. 67:504-512(1999).  
CC -1- FUNCTION: PROBABLY FACILITATES NICKEL INCORPORATION.  
CC -1- SIMILARITY: BELONGS TO THE URED FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF056321; AAD13736.1; -  
DR EMBL; AF048781; AAD13726.1; -  
DR InterPro; IPR002669; Ured.  
DR Pfam; PF01774; Ured; 1.  
KW Nickel.  
SQ SEQUENCE 271 AA; 29247 MW; 0042A71CC3F00684 CRC64;

Query Match 42.9%; Score 39; DB 1; Length 271;  
Best Local Similarity 54.5%; Pred. No. 17;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 VSYRGISYRRS 16  
: : : : : | | | | :  
DB 112 IAYRGASYRQT 122

RESULT 13  
BCAL\_RAT  
ID BCAL\_RAT STANDARD; PRT; 968 AA.  
AC Q63767; Q63766;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CRK-ASSOCIATED SUBSTRATE (P130CAS) (BREAST CANCER ANTI-ESTROGEN  
RESISTANCE 1 PROTEIN).  
GN BCAR1 OR CRKAS OR CAS.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Fibroblast;  
RX MEDLINE=94349922; PubMed=8070403;  
RA Sakai R., Iwamatsu A., Hirano N., Ogawa S., Tanaka T., Mano H.,  
Yazaki Y., Hirai H.;  
RT "A novel signaling molecule, pl30, forms stable complexes in vivo with  
v-Crk and v-Src in a tyrosine phosphorylation-dependent manner.";  
RL EMBO J. 13:3748-3756(1994).  
RN [2]  
RP TYROSINE PHOSPHORYLATION BY FOCAL ADHESION KINASE.  
RX MEDLINE=98030586; PubMed=9360983;  
RA Tachibana K., Urano T., Fujita H., Ohashi Y., Kamiguchi K., Iwata S.,  
Hirai H., Morimoto C.;  
RT "Tyrosine phosphorylation of Crk-associated substrates by focal

adhesion kinase. A putative mechanism for the integrin-mediated  
tyrosine phosphorylation of crk-associated substrates.";  
J. Biol. Chem. 272:29083-29090(1997).  
- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE  
FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION.  
IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY).  
SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1,  
ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL.  
SUBCELLULAR LOCATION: FOCAL ADHESIONS AND STRESS FIBERS.  
UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO  
THE MEMBRANE UPON TYROSINE PHOSPHORYLATION.  
ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)  
AND A SHORT ISOFORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING,  
TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHER EXPRESSION IN LUNG,  
INTESTINE AND TESTIS.  
DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING  
MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN  
CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-  
BINDING SITES PUTATIVELY BIND CRK, NCK AND ABL SH2 DOMAINS. THE  
HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOHYPHAL  
GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.  
DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM  
RESPONSE ELEMENT (SRE).  
PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE  
DYVHL MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE  
PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE  
RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN  
MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX.  
DISEASE: APPEARS TO HAVE A CENTRAL FUNCTION IN TRANSFORMATION OF  
SOME CELL TYPES.  
SIMILARITY: CONTAINS 1 SH3 DOMAIN. NECESSARY FOR THE LOCALISATION  
OF THE PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-  
RICH REGION OF FOCAL ADHESION KINASE 1.  
SIMILARITY: CONTAINS 1 SH3-BINDING DOMAIN. THE SH3 BINDING DOMAIN  
BINDS TO THE SRC SH3 DOMAIN.  
SIMILARITY: BELONGS TO THE CAS FAMILY.  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; D29766; BAA06169.1; -;  
EMBL; D29766; BAA06170.1; -;  
HSP; P29354; IGR1  
InterPro: IPR001452; SH3.  
Pfam: PF00018; SH3; 1.  
PRINTS: PR00452; SH3DOMAIN.  
SMART: SM00326; SH3; 1.  
PROSITE: PS50002; SH3; 1.  
Phosphorylation: SH3 domain; SH3-binding; Cell adhesion;  
Alternative splicing  
DOMAIN 97 159  
PRO-RICH.  
DOMAIN 168 181  
SUBSTRATE FOR KINASES.  
DOMAIN 213 514  
SER-RICH.  
DOMAIN 520 712  
SH3-BINDING (POTENTIAL).  
DOMAIN 733 741  
DIVERGENT HELIX-LOOP-HELIX MOTIF.  
DOMAIN 844 894  
MISSING (IN SHORT ISOFORM).  
VARSPLIC 5 98  
SEQUENCE 968 AA; 104262 MW; E861641BF6D8D377 CRC64;

Query Match 42.9%; Score 39; DB 1; Length 968;  
Best Local Similarity 56.2%; Pred. No. 61;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
QY 2 WSPRVSYRGISYRRSR 17  
| | | | | | | | | |  
Db 23 WSPRVSRPQSVRAAR 38

RESULT 14  
L2CC\_DROME STANDARD; PRT; 203 AA.  
AC P24156;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE L(2)37CC PROTEIN.  
GN L(2)37CC.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=86312887; PubMed=3092183;  
RX Eveleth D.D. Jr., Marsh J.L.;  
RA "Sequence and expression of the Cc gene, a member of the dopa  
decarboxylase gene cluster of Drosophila: possible translational  
regulation.";  
RT Nucleic Acids Res. 14:6169-6183(1986).  
RL -1- FUNCTION: REQUIRED FOR LARVAL METABOLISM OR FOR THE PROGRESSION  
OF THE LARVA INTO A PUPA.  
CC -1- SIMILARITY: BELONGS TO THE PROHIBITIN FAMILY.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; X04228; CAA27810.1; -;  
EMBL; X04227; CAA27807.1; -;  
PIR: C25511; C25511.  
FlyBase; FBgn002031; l(2)37Cc.  
InterPro: IPR001107; Band\_7.  
DR Pfam; PF01145; Band\_7; 1.  
DR SMART; SM00244; PHB; 1.  
KW Developmental protein.  
KW SEQUENCE 203 AA; 22817 MW; B15D085CC0862A11 CRC64;  
Query Match 41.8%; Score 38; DB 1; Length 203;  
Best Local Similarity 50.0%; Pred. No. 18;  
Matches 10; Conservative 1; Mismatches 5; Indels 4; Gaps 1;  
QY 2 WSPRV---SYRGISYRRSR 17  
| | | | | | | | | |  
Db 158 WSLRLIDRPRYLTSPRSR 177  
RESULT 15  
HB22\_HUMAN STANDARD; PRT; 261 AA.  
ID HB22\_HUMAN  
AC P01919.  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(WI.1) BETA CHAIN  
DE PRECURSOR (DOB1\*0501).  
GN HLA-DQB1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=86055719; PubMed=2998758;  
RX Tonnelie C., Demars R., Long E.O.;  
RA



RT "DQ beta: a new beta chain gene in HLA-D with a distinct regulation  
of expression.";  
RL EMBO J. 4:2839-2847(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88006310; PubMed=2888727;  
RA Turco E., Care A., Compagnone-Post P., Robinson C., Cascino I.,  
RN Trucco M.;  
RT "Allelic forms of the alpha- and beta-chain genes encoding  
DQ1-positive heterodimers.";  
RL Immunogenetics 26:282-290(1987).  
RN [3]  
RP SEQUENCE OF 33-261 FROM N.A. (CLONE PII-BETA-2).  
RX MEDLINE=84031733; PubMed=6415003;  
RA Larhammar D., Andersson G., Andersson M., Bill P., Boehme J.,  
RN Claesson L., Denaro M., Emmoth E., Gustafsson K., Hammarling U.,  
RA Heldin E., Hyldig-Nielsen J.J., Lind P., Schenning L., Servenius B.,  
RN Widmark E., Rask L., Peterson P.A.;  
RT "Molecular analysis of human class II transplantation antigens and  
their genes.";  
RL Hum. Immunol. 8:95-103(1983).  
RN [4]  
RP SEQUENCE OF 33-261 FROM N.A.  
RX MEDLINE=97083137; PubMed=8929711;  
RA Yasunaga S., Kimura A., Hamaguchi K., Ronningen K.S., Sasazuki T.;  
RN "Different contribution of HLA-DR and -DQ genes in susceptibility and  
resistance to insulin-dependent diabetes mellitus (IDDM).";  
RL Tissue Antigens 47:37-48(1996).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X03068; CAA26872.1; -;  
DR EMBL: M17564; AAA59765.1; -;  
DR EMBL: L34101; AAC41969.1; -;  
DR PIR: A02232; HLHU2C.  
DR PIR: C24669; C24669.  
DR HSSP: P13760; 2SEB.  
DR MIN: 604305; -;  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003597; Ig\_cl.  
DR InterPro: IPR000353; MHC\_II\_beta.  
DR Pfam: PF00047; Ig; 1.  
DR ProDom: PD000328; MHC\_II\_beta; 1.  
DR SMART: SM00407; IGL; 1.  
DR PROSITE: PS00290; IG\_MHC; 1.  
KW MHC II; Transmembrane; Glycoprotein; Signal.  
FT SIGNAL 1 32  
FT CHAIN 33 261  
FT FT  
FT HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,  
FT DQ(W1.1) BETA CHAIN.  
FT DOMAIN 33 126  
FT DOMAIN 127 220  
FT DOMAIN 221 230  
FT DOMAIN 231 251  
FT TRANSMEM 252 261  
FT DOMAIN 47 111  
FT DISULFID 149 205  
FT DISULFID 51 51  
FT CARBOHYD 261 AA; 29748 MW; 2F5D8FDC413D1BA5 CRC64;  
SQ SEQUENCE

Query Match 41.8%; Score 38; DB 1; Length 261;  
Best Local Similarity 53.8%; Pred. No. 24;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 3 SFRVSYRGISYRR 15  
:: |:|||| |

Db 114 NYEVAYRGILQRR 126  
Search completed: February 12, 2002, 12:39:53  
Job time: 306 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:40 ; Search time 232.64 Seconds  
(without alignments)  
10.689 Million cell updates/sec

Title: US-09-485-571-26  
Perfect score: 91  
Sequence: 1 KWSFRVSYRGISYR 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_17.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	48	52.7	570	10 Q9LQE3	Q9LQE3 arabidopsis
2	47	51.6	615	10 Q9C7I9	Q9C7I9 arabidopsis
3	44.5	48.9	88	2 Q9KMK7	Q9KMK7 vibrio chol
4	44	48.4	307	1 Q9HSY4	Q9HSY4 halobacteri
5	44	48.4	378	1 Q9HSZ6	Q9HSZ6 halobacteri
6	44	48.4	620	10 Q9LWK2	Q9LWK2 arabidopsis
7	44	48.4	620	10 Q9C8N9	Q9C8N9 arabidopsis
8	43	47.3	660	5 Q17248	Q17248 boophilus m
9	43	47.3	781	5 Q9NF24	Q9NF24 caenorhabdi
10	43	47.3	1106	13 Q42291	Q42291 gallus gall
11	42	46.2	182	2 Q9RBT9	Q9RBT9 burkholderi
12	42	46.2	318	2 Q53813	Q53813 shigella bo
13	42	46.2	480	2 Q9RSL6	Q9RSL6 deinococcus
14	41	45.1	279	10 Q80901	Q80901 arabidopsis
15	41	45.1	306	11 Q9JUU7	Q9JUU7 rattus norv
16	41	45.1	323	5 Q9VAQ8	Q9VAQ8 drosophila
17	41	45.1	347	11 Q9WUX9	Q9WUX9 rattus norv
18	41	45.1	366	7 P79587	P79587 rattus norv
19	41	45.1	378	2 Q9AJU2	Q9AJU2 streptomyce

20	41	45.1	770	5 Q20908	Q20908 caenorhabdi
21	40	44.0	52	9 Q9MBU4	Q9MBU4 chlamydia p
22	40	44.0	114	11 Q9DOK5	Q9DOK5 mus musculu
23	40	44.0	146	2 Q9X5Q7	Q9X5Q7 streptomyce
24	40	44.0	330	1 Q9YBP0	Q9YBP0 aeropyrum p
25	40	44.0	397	2 Q85368	Q85368 enterococcu
26	40	44.0	411	4 Q43916	Q43916 homo sapien
27	40	44.0	411	11 Q9EQC0	Q9EQC0 mus musculu
28	40	44.0	583	3 Q9C2K0	Q9C2K0 neurospora
29	40	44.0	713	2 Q9FC87	Q9FC87 streptomyce
30	40	44.0	901	2 Q9PH65	Q9PH65 xylella fas
31	40	44.0	1081	11 P97423	P97423 mus musculu
32	40	44.0	1124	4 Q9Y6S2	Q9Y6S2 homo sapien
33	40	44.0	1601	5 Q9N4C2	Q9N4C2 caenorhabdi
34	39.5	43.4	124	2 Q9AD90	Q9AD90 streptomyce
35	39.5	43.4	781	10 Q9LWJ6	Q9LWJ6 arabidopsis
36	39.5	43.4	846	5 Q17897	Q17897 caenorhabdi
37	39	42.9	242	1 Q9Y8W0	Q9Y8W0 aeropyrum p
38	39	42.9	255	10 Q9FMA0	Q9FMA0 arabidopsis
39	39	42.9	597	1 Q58537	Q58537 pyrococcus
40	39	42.9	767	10 Q9LQE8	Q9LQE8 arabidopsis
41	39	42.9	901	10 Q9ZPF8	Q9ZPF8 arabidopsis
42	39	42.9	1124	4 Q99699	Q99699 homo sapien
43	39	42.9	1148	10 Q9C6N3	Q9C6N3 arabidopsis
44	38.5	42.3	563	2 Q05770	Q05770 mycobacteri
45	38.5	42.3	859	13 Q98UH9	Q98UH9 oryzias lat

ALIGNMENTS

RESULT 1

ID Q9LQE3 PRELIMINARY; PRT; 570 AA.

AC Q9LQE3; 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE F1504.42.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,

RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chioi J., Choi E.,

RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,

RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,

RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,

RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,

RA Theologis A., Ecker J.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC007887; AAF79360.1; -

DR InterPro: IPR003340; B3.

DR Pfam: PF02362; B3; 1.

SQ SEQUENCE 570 AA; 65041 MW; E64A987BA713DF7E CRC64;

Query Match 52.7%; Score 48; DB 10; Length 570;  
Best Local Similarity 64.38; Pred. No. 7.6;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYR 14

```
Db 175 QMSFRVSYRGTPQR 188
      :||||| ||||| |
RESULT 2
Q9C7I9 PRELIMINARY; PRT; 615 AA.
AC Q9C7I9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 69.8 KDA PROTEIN.
GN T911.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
DR EMBL: AC069160; AAG51458.1; -.
DR InterPro: IPR003311; AUX_1AA.
DR InterPro: IPR003340; B3.
DR Pfam: PF02309; AUX_1AA; 1.
DR Pfam: PF02362; B3; 1.
KW Hypothetical protein.
SQ SEQUENCE 615 AA; 69755 MW; 817E7D03190622F7 CRC64;

Query Match 51.6%; Score 47; DB 10; Length 615;
Best Local Similarity 56.2%; Pred. No. 12;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRS 16
      :| | | | | | | | | |
Db 175 QMRFRHSYRGTPQRS 190
      :| | | | | | | | | |

RESULT 3
Q9KMK7 PRELIMINARY; PRT; 88 AA.
AC Q9KMK7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN VCA0332.
GN VCA0332.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.

STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Colishmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL: AE004371; AAF96240.1; -.
DR TIGR: VCA0332; -.
KW Complete proteome.
SQ SEQUENCE 88 AA; 10092 MW; B173D34A34B6A870 CRC64;

Query Match 48.9%; Score 44.5; DB 2; Length 88;
Best Local Similarity 52.4%; Pred. No. 3.8;
Matches 11; Conservative 2; Mismatches 3; Indels 5; Gaps 1;

QY 2 WSPFRVSYRG-----ISYRRSR 17
      || :||| || |||||
Db 58 WSGVITRGTNIRIISYRRSR 78
      || :||| || |||||

RESULT 4
Q9HSY4 PRELIMINARY; PRT; 307 AA.
AC Q9HSY4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VNG0026C.
GN VNG0026C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir-D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W., Dale H.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Spudich J.L., Jung K.-H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Omer A.D.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Dassarma S.;
RA Eberhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE004971; AAG18666.1; -.
KW Complete proteome.
SQ SEQUENCE 307 AA; 34492 MW; AB92A26FF80AF39 CRC64;

Query Match 48.4%; Score 44; DB 1; Length 307;
Best Local Similarity 42.9%; Pred. No. 18;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYR 14
      ||||| | | | |
Db 181 KWAFTLYEQVAYK 194
      ||||| | | | |

RESULT 5
Q9HSZ6 PRELIMINARY; PRT; 378 AA.
AC Q9HSZ6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
```



Best Local Similarity 41.2%; Pred. No. 74; Indels 0; Gaps 0;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 KWSFRVSYRGISYRRSR 17  
DB 388 KWPIDISYRDSEHRRKK 404

RESULT 10  
042291 PRELIMINARY; PRT; 1106 AA.

ID O42291  
AC O42291;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE JANUS TYROSINE KINASE.  
GN JAK  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SPLEEN;  
RA Sofer L., Kampa D., Burnside J.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF034576; AAC34195.1; -.  
DR HSSP; P12931; 1FMK.  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR000980; SH2.  
DR InterPro; IPR001245; Tyr\_kin.  
DR Pfam; PF00009; pkinase; 2.  
DR Pfam; PF00017; SH2; 1.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00252; SH2; 1.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 2.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS00001; SH2; 1.  
DR ATP-binding; Transferase; Tyrosine-protein kinase.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1106 AA; 124750 MW; 004F75F851A282B1 CRC64;

Query Match 47.3%; Score 43; DB 13; Length 1106;  
Best Local Similarity 57.1%; Pred. No. 1.le+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 FRVSYRGISYRRSR 17  
DB 881 FIVKRGVCYSRGR 894

RESULT 11  
Q9RBT9 PRELIMINARY; PRT; 182 AA.

ID Q9RBT9  
AC Q9RBT9;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE P-HYDROXYLAMINOENZOATE LYASE.  
GN PNB.  
OS Burkholderia pickettii (Pseudomonas pickettii).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YH105;  
RA Yabannavar A., Zylstra G.J.;  
RT \*Analysis of the genes for p-nitrobenzoate degradation from Ralstonia

RESULT 8  
Q17248 PRELIMINARY; PRT; 660 AA.  
AC Q17248;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE ANGIOTENSIN-CONVERTING ENZYME-LIKE PROTEIN PRECURSOR.  
GN BM91.  
OS Boophilus microplus (Cattle tick).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.  
OX NCBI\_TaxID=6941;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=WHOLE TICKS;  
RA Whitfield P.L., Johnson M.C., Smith D.R.J., Snelson V.J., Clifton G.F.,  
RA Brown G.S., Cairns D., Foy A.B., Irving D.O.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U62809; AAB04998.1; -.  
DR InterPro; IPR001548; Peptidase\_M2.  
DR Pfam; PF01401; Peptidase\_M2; 1.  
DR PRINTS; PR00791; PEPTIDASEA.  
DR ProDom; PD004184; Peptidase\_M2; 1.  
KW Signal.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 660 POTENTIAL.  
SQ SEQUENCE 660 AA; 75257 MW; 6F164CF70C938E63 CRC64;

Query Match 47.3%; Score 43; DB 5; Length 660;  
Best Local Similarity 45.5%; Pred. No. 61;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 WSRVSYRGIS 12  
DB 488 WEYRIKQGVGS 498

RESULT 9  
Q9NF24 PRELIMINARY; PRT; 781 AA.  
AC Q9NF24;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE Y10588A.A PROTEIN.  
GN Y10588A.A.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sulston J.E.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RX MEDLINE-99069613; PubMed-9851916;  
RA none;  
RT "Genome sequencing of the nematode C.elegans: A platform for investigating biology";  
RL Science 282:2012-2018(1998).  
DR EMBL; AL132876; CAB00841.1; -.  
DR InterPro; IPR001930; Aladiptase.  
DR InterPro; IPR000130; Zn\_Mtpeptidse.  
DR Pfam; PF01433; Peptidase\_M1; 1.  
DR PRINTS; PR00756; ALADIPTASE.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN 1.  
SQ SEQUENCE 781 AA; 91124 MW; 1E3FC20D8E071022 CRC64;

Query Match 47.3%; Score 43; DB 5; Length 781;

RT pickettii VH105.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF187879; AAF01444.1; -;  
KW Lyase.  
SQ SEQUENCE 182 AA; 20053 MW; 45FA1636CAC43E06 CRC64;

Query Match 46.2%; Score 42; DB 2; Length 182;  
Best Local Similarity 43.8%; Pred. No. 22;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 WSRVSYRGISYRRSR 17  
| : : | : |||||  
Db 63 WAANIDIEGPNYRRSR 78

RESULT 12  
Q53813 PRELIMINARY; PRT; 318 AA.  
ID Q53813  
AC Q53813  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
DE BV' GENE PRODUCT (FRAGMENT).  
OS Shigella boydii.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=621;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91286192; PubMed=2061288;  
RA Tomimaga A., Ikemizu S., Enomoto M.;  
RT "Site-specific recombinase genes in three Shigella subgroups and  
RT nucleotide sequences of a pinB gene and an invertible B segment from  
RT Shigella boydii.";  
RL J. Bacteriol. 173:4079-4087(1991).  
DR EMBL: D00660; BAA00555.1; -;  
FT NON\_TER 318 318  
SQ SEQUENCE 318 AA; 34335 MW; 93CC950314D4E275 CRC64;

Query Match 46.2%; Score 42; DB 2; Length 318;  
Best Local Similarity 62.5%; Pred. No. 41;  
Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 4 FRVSYR--GISYRRSR 17  
|||:| | | : | : |  
Db 205 FRVYRNGGIFYRSAR 220

RESULT 13  
Q9RSL6 PRELIMINARY; PRT; 480 AA.  
ID Q9RSL6  
AC Q9RSL6  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE MOLYBDATE METABOLISM REGULATOR-RELATED PROTEIN.  
GN DR2108.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL: AE002046; AAF11657.1; -;  
DR TIGR: DR2108; -;  
KW Complete proteome.  
SQ SEQUENCE 480 AA; 52430 MW; B67C9175AAC14281 CRC64;

Query Match 46.2%; Score 42; DB 2; Length 480;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 WSRVSYRGISYRRSR 17  
| | | | : | : | : |  
Db 234 WMYRVDARGVYGHRSR 249

RESULT 14  
O80901 PRELIMINARY; PRT; 279 AA.  
ID O80901  
AC O80901  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 31.8 KDA PROTEIN T6A23.30.  
GN T6A23.30.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
RA Shen M., Renning C.M., Fraser C.M., Somerville C.R., Venter J.C.;  
RT "Arabidopsis thaliana chromosome II BAC T6A23 genomic sequence.";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC005499; AAC67368.1; -;  
SQ SEQUENCE 279 AA; 31775 MW; 393AAE4AE046B8B2 CRC64;

Query Match 45.1%; Score 41; DB 10; Length 279;  
Best Local Similarity 80.0%; Pred. No. 52;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 SFRVSYRGIS 12  
: ||||| | |  
Db 99 AFRVSYHGHS 108

RESULT 15  
Q9JJU7 PRELIMINARY; PRT; 306 AA.  
ID Q9JJU7  
AC Q9JJU7  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE RT1-EL PROTEIN (FRAGMENT).  
GN RT1-EL.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LEWIS; TISSUE=BRAIN STRIATUM;  
RA McLaren F.H.;  
RL Thesis (2000), Laboratory of Functional Immunogenetics,  
RL The Babraham Institute, Cambridge, U.K.  
CC -I- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE  
CC IMMUNE SYSTEM (BY SIMILARITY).

Wed Feb 13 07:52:24 2002

CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN) (BY SIMILARITY).  
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
CC DOMAIN.

DR EMBL; AJ276126; CAB86228.2; -.  
DR InterPro; IPR003597; Ig\_cl.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR001039; MHC\_I.  
DR Pfam; PF00047; Ig; 1.  
DR Pfam; PF00129; MHC\_I; 1.  
DR PRODOM; PD000050; MHC\_I; 1.  
DR SMART; SM00407; IGcl; 1.  
DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.  
KW Glycoprotein; Transmembrane.  
FT NON\_TER 1  
SQ SEQUENCE 306 AA; 35168 MW; 26BA7F9E4B960F23 CRC64;

Query Match 45.18; Score 41; DB 11; Length 306;  
Best Local Similarity 63.6%; Pred. No. 57;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KWSRVSVYRGI 11  
Db 71 EWSFRVSLRNL 81

Search completed: February 12, 2002, 12:38:41  
Job time: 754 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:33 ; Search time 242.57 Seconds  
(without alignments)  
5.191 Million cell updates/sec

Title: US-09-485-571-27

Perfect score: 91

Sequence: 1 RWSFRVSYRGISYRRSR 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	91	100.0	17	20 AAW99414	Tachyplesin deriva
2	88	96.7	17	20 AAW99413	Tachyplesin deriva
3	88	96.7	17	21 AAY93617	Peptide which may
4	76	83.5	17	16 AAR75806	Antimicrobial tach
5	76	83.5	17	21 AAY69610	Tachyplesin analog
6	72	79.1	17	16 AAR75819	Antimicrobial tach
7	72	79.1	17	21 AAY69609	Generic tachyplesin
8	72	79.1	17	21 AAY69617	Tachyplesin analog
9	71	78.0	17	16 AAR75807	Antimicrobial tach
10	71	78.0	17	16 AAR75808	Antimicrobial tach
11	71	78.0	21	16 AAR75816	Antimicrobial tach

12	71	78.0	35	16	AAR75810	Antimicrobial tach
13	70	76.9	17	16	AAR75822	Antimicrobial tach
14	68	74.7	17	10	AA91671	New lipopolysaccha
15	68	74.7	17	11	AAR06266	Antiviral peptide
16	68	74.7	17	11	AAR06861	Tachyplesin I. Li
17	68	74.7	17	11	AAR06862	Tachyplesin II. L
18	68	74.7	17	13	AAR23112	Bacterial shock tr
19	68	74.7	17	13	AAR23113	Bacterial shock tr
20	68	74.7	17	14	AAR38490	Tachyplesin-II. Ta
21	68	74.7	17	14	AAR38489	Tachyplesin-I. Tac
22	68	74.7	17	16	AAR75805	Tachyplesin, an an
23	68	74.7	17	19	AAW66465	Cationic peptide t
24	68	74.7	17	19	AAW66466	Cationic peptide t
25	68	74.7	17	21	AAY91764	Cationic peptide t
26	68	74.7	17	21	AAY91765	Cationic peptide t
27	68	74.7	17	21	AAY69608	Tachyplesin (TP),
28	68	74.7	17	21	AAY69614	Tachyplesin peptid
29	68	74.7	17	22	AAR91394	Tachyplesin peptid
30	68	74.7	39	16	AAR75817	Antimicrobial tach
31	67	73.6	17	16	AAR75815	Antimicrobial tach
32	67	73.6	17	16	AAR75803	Antimicrobial tach
33	65	71.4	17	11	AAR08202	Gigaslin II. Tachy
34	65	71.4	17	13	AAR23114	Bacterial shock tr
35	65	71.4	17	14	AAR38491	Tachyplesin-III. T
36	64	70.3	17	16	AAR75820	Antimicrobial tach
37	64	70.3	17	16	AAR75813	Antimicrobial tach
38	64	70.3	17	16	AAR75814	Antimicrobial tach
39	64	70.3	17	21	AAY69611	Tachyplesin analog
40	64	70.3	17	21	AAY69612	Tachyplesin analog
41	64	70.3	17	21	AAY69613	Tachyplesin analog
42	64	70.3	17	21	AAY69615	Tachyplesin analog
43	64	70.3	17	21	AAY69616	Tachyplesin analog
44	64	70.3	18	11	AAR03670	Polypheumusin I. L
45	64	70.3	18	11	AAR06863	Polypheumusin I. L

#### ALIGNMENTS

RESULT 1  
AAW99414  
ID AAW99414 standard; peptide; 17 AA.  
XX AAW99414;  
AC AAW99414;  
XX  
DT 08-JUN-1999 (first entry)  
XX  
DE Tachyplesin derivative peptide SM2307.  
XX  
KW Linear; tachyplesin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.  
XX  
OS Synthetic.  
XX  
PN WO9907728-A2.  
XX  
PD 18-FEB-1999.  
XX  
PF 06-AUG-1998; 98WO-FR01757.  
XX  
PR 12-AUG-1997; 97FR-0010297.  
XX  
PA (SYNT-) SYNT:EM SA.  
XX  
PI Calas B, Chavanieu A, Grassy G, Kaczorek M;  
XX WPI; 1999-190034/16.  
XX  
PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
XX as carriers to deliver active agents into cells

RESULT 4

```

t
AAR75806
ID AAR75806 standard; peptide; 17 AA.
XX
AC AAR75806;
XX
DT 07-FEB-1996 (first entry)
XX
DE Antimicrobial tachyplesin peptide derivative.
XX
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;
KW infection.
XX
OS Synthetic.
XX
PN WO9516776-A1.
XX
PD 22-JUN-1995.
XX
PF 19-DEC-1994; 94WO-US14619.
XX
PR 17-DEC-1993; 93US-0168809.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Putman RJ, Rao AG, Rao A;
XX
DR WPI; 1995-231570/30.
XX
PT New peptide derivs. of tachyplesin - having antimicrobial activity,
PT used against plant pathogenic fungi or human or animal infections
XX
PS Claim 1; Page 29; 45pp; English.
XX
CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin
CC (AAR75805) a small peptide isolated from Japanese horseshoe crab
CC haemocytes which has antimicrobial properties. The peptide derivatives
CC also have antimicrobial activity and can be used for killing and
CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.
CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,
CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can
CC also be used for treating and preventing infection in humans and
CC animals.
XX
SQ Sequence 17 AA;

Query Match 83.5%; Score 76; DB 16; Length 17;
Best Local Similarity 70.8%; Pred. NO. 5.4e-06;
Matches 12; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17
Db :|||:||||:||||:
1 kwafvayrgiayrrar 17

RESULT 5
AAY69610
ID AAY69610 standard; peptide; 17 AA.
XX
AC AAY69610;
XX
DT 08-MAY-2000 (first entry)
XX
DE Tachyplesin analogue, TPA.
XX
KW Tachyplesin analogue; generic; antimicrobial; disulphide bond;
KW antifungal; antiviral; antimicrobial; transgenic plant.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 17 /note= "C-terminal amide"

```

```

XX US60:5941-A.
XX PN
XX 18-JAN-2000.
XX PD
XX 31-OCT-1997; 97US-0962034.
XX PF
XX 31-OCT-1997; 97US-0962034.
XX PR
XX (PION-) PIONEER HI-BRED INT INC.
XX PA
XX Rao AG;
XX PI
XX WPI; 2000-126327/11.
XX DR
XX New tachyplesin analogs useful for controlling fungal and bacterial
XX activity in agricultural and medical applications and for controlling
XX plant viruses have four cysteine substitutions -
XX Example 1; Page -: 17pp; English.
XX PS
XX Sequences AAY69610-Y69611 and AAY69617 represent tachyplesin analogues
XX used in an exemplification of the present invention, in which the
XX native tachyplesin cysteine residues are replaced with Ala, Leu and
XX Asp, respectively. Tachyplesin (AAY69608) is a naturally occurring
XX antimicrobial peptide which contains two disulphide bonds which help
XX to maintain its tertiary structure. The invention relates to novel
XX peptide analogues of tachyplesin (Y696912-AAY69614, AAY69616) in which
XX the cysteine residues at positions 3, 7, 12 and 16 of the native
XX tachyplesin are replaced by the hydrophobic amino acids isoleucine,
XX valine, methionine, phenylalanine or tyrosine, the same amino acid being
XX present at all four positions. Despite being unable to form
XX intramolecular disulphide bonds, the analogues are functional as
XX antimicrobial agents. The tachyplesin analogues are useful for
XX controlling fungal and viral activity in agricultural and medical
XX applications and for controlling plant viruses. They can also be
XX expressed in transgenic plants, preferably wheat, sorghum, sunflower,
XX soya or especially maize plants to provide resistance to pathogenic fungi
XX and viruses. Note: The present sequence is not shown in the
XX specification, but is derived from the generic tachyplesin analogue
XX sequence given in column 23.
XX SQ Sequence 17 AA;

Query Match 83.5%; Score 76; DB 21; Length 17;
Best Local Similarity 70.8%; Pred. NO. 5.4e-06;
Matches 12; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17
Db :|||:||||:||||:
1 kwafvayrgiayrrar 17

RESULT 6
AAR75819
ID AAR75819 standard; peptide; 17 AA.
XX
AC AAR75819;
XX
DT 07-FEB-1996 (first entry)
XX
DE Antimicrobial tachyplesin peptide derivative.
XX
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;
KW infection.
XX
OS Synthetic.
XX
PN WO9516776-A1.
XX
PD 22-JUN-1995.
XX

```

PF 19-DEC-1994; 94WO-US14619.  
XX  
PR 17-DEC-1993; 93US-0168809.  
XX  
XX (PION-) PIONEER HI-BRED INT INC.  
PA  
PI Putman RJ, Rao AG, Rao A;  
XX  
XX WPI; 1995-231570/30.  
DR  
XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
PT used against plant pathogenic fungi or human or animal infections  
PT  
XX Claim 1; Page 35; 45pp; English.  
PS  
XX AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
CC haemocytes which has antimicrobial properties. The peptide derivatives  
CC also have antimicrobial activity and can be used for killing and  
CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
CC also be used for treating and preventing infection in humans and  
CC animals.  
XX  
XX Sequence 17 AA;  
SQ  
Query Match 79.1%; Score 72; DB 16; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.4e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 RWSFRVSYRGISYRRSR 17  
Db : l l l l l l l l l l  
1 kwkfrvkgikyrkr 17  
RESULT 7  
AAY69609  
ID AAY69609 standard; peptide; 17 AA.  
XX  
XX AAY69609;  
AC  
XX 08-MAY-2000 (first entry)  
DT  
XX Generic tachyplesin (TP) analogue antimicrobial peptide.  
DE  
XX Tachyplesin analogue; generic; antimicrobial; disulphide bond;  
KW antifungal; antiviral; antimicrobial; transgenic plant.  
KW  
XX Synthetic.  
OS  
XX Key Location/Qualifiers  
FH Misc-difference 3 /label= Ile, Val, Met, Phe, Tyr  
FT  
FT Misc-difference 7 /label= Ile, Val, Met, Phe, Tyr  
FT  
FT Misc-difference 12 /label= Ile, Val, Met, Phe, Tyr  
FT  
FT Misc-difference 16 /label= Ile, Val, Met, Phe, Tyr  
FT  
FT /label= Ile, Val, Met, Phe, Tyr  
FT /note= "The molecule has the same amino acid at all four  
FT of the above positions"  
FT Modified-site 17 /note= "C-terminal amide"  
FT  
XX US6015941-A.  
XX  
XX 18-JAN-2000.  
PD  
XX 31-OCT-1997; 97US-0962034.  
XX  
XX 31-OCT-1997; 97US-0962034.  
PR

XX (PION-) PIONEER HI-BRED INT INC.  
PA  
XX Rao AG;  
XX  
XX WPI; 2000-126327/11.  
DR  
XX New tachyplesin analogs useful for controlling fungal and bacterial  
PT activity in agricultural and medical applications and for controlling  
PT plant viruses have four cysteine substitutions -  
PT  
XX Claim 1; Column 23; 17pp; English.  
PS  
XX This sequence represents a generic tachyplesin (TP) analogue which has  
CC antimicrobial activity. Tachyplesin (AAY69608) is a naturally occurring  
CC antimicrobial peptide which contains two disulphide bonds which help  
CC to maintain its tertiary structure. The invention relates to novel  
CC peptide analogues of tachyplesin (Y69612-AAY69614, AAY69616) in which  
CC the cysteine residues at positions 3, 7, 12 and 16 of the native  
CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,  
CC valine, methionine, phenylalanine or tyrosine, the same amino acid being  
CC present at all four positions. Despite being unable to form  
CC intramolecular disulphide bonds, the analogues are functional as  
CC antimicrobial agents. The tachyplesin analogues are useful for  
CC controlling fungal and viral activity in agricultural and medical  
CC applications and for controlling plant viruses. They can also be  
CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,  
CC soya or especially maize plants to provide resistance to pathogenic fungi  
CC and viruses.  
XX  
XX Sequence 17 AA;  
SQ  
Query Match 79.1%; Score 72; DB 21; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.4e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 RWSFRVSYRGISYRRSR 17  
Db : l l l l l l l l l l  
1 kwkfrvxyrgixyrxr 17  
RESULT 8  
AAY69617  
ID AAY69617 standard; peptide; 17 AA.  
XX  
XX AAY69617;  
AC  
XX 08-MAY-2000 (first entry)  
DT  
XX Tachyplesin analogue, TPD.  
DE  
XX Tachyplesin analogue; generic; antimicrobial; disulphide bond;  
KW antifungal; antiviral; antimicrobial; transgenic plant.  
KW  
XX Synthetic.  
OS  
XX Key Location/Qualifiers  
FH Modified-site 17 /note= "C-terminal amide"  
FT  
FT US6015941-A.  
XX  
XX 18-JAN-2000.  
PD  
XX 31-OCT-1997; 97US-0962034.  
XX  
XX 31-OCT-1997; 97US-0962034.  
PR  
XX (PION-) PIONEER HI-BRED INT INC.  
PA  
XX Rao AG;  
PI  
XX

DR WPI; 2000-126327/11.  
 XX  
 PT New tachyplesin analogs useful for controlling fungal and bacterial  
 PT activity in agricultural and medical applications and for controlling  
 PT plant viruses have four cysteine substitutions -  
 XX  
 XX Example 1; Page -: 17pp; English.  
 PS  
 XX Sequences AAY69610-Y69611 and AAY69617 represent tachyplesin analogues  
 CC used in an exemplification of the present invention, in which the  
 CC native tachyplesin cysteine residues are replaced with Ala, Leu and  
 CC Asp, respectively. Tachyplesin (AAY69608) is a naturally occurring  
 CC antimicrobial peptide which contains two disulphide bonds which help  
 CC to maintain its tertiary structure. The invention relates to novel  
 CC peptide analogues of tachyplesin (Y696912-AAY69614, AAY69616) in which  
 CC the cysteine residues at positions 3, 7, 12 and 16 of the native  
 CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,  
 CC valine, methionine, phenylalanine or tyrosine, the same amino acid being  
 CC present at all four positions. Despite being unable to form  
 CC intramolecular disulphide bonds, the analogues are functional as  
 CC antimicrobial agents. The tachyplesin analogues are useful for  
 CC controlling fungal and viral activity in agricultural and medical  
 CC applications and for controlling plant viruses. They can also be  
 CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,  
 CC soya or especially maize plants to provide resistance to pathogenic fungi  
 CC and viruses. Note: The present sequence is not shown in the  
 CC specification, but is derived from the generic tachyplesin analogue  
 XX sequence given in column 23.  
 XX  
 XX Sequence 17 AA;

Query Match 79.1%; Score 72; DB 21; Length 17;  
 Best Local Similarity 70.6%; Pred. No. 2.4e-05;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
 : | ||| |||| ||| |  
 Db 1 kwdfvdyrgdyrrdr 17

RESULT 9  
 AAR75807  
 ID AAR75807 standard; peptide; 17 AA.  
 AC  
 AC AAR75807;  
 XX  
 XX 07-FEB-1996 (first entry)  
 DT  
 XX Antimicrobial tachyplesin peptide derivative.  
 DE  
 XX Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 KW infection.  
 KW  
 XX Synthetic.  
 OS  
 XX WO9516776-A1.  
 PN  
 XX 22-JUN-1995.  
 PD  
 XX 19-DEC-1994; 94WO-US14619.  
 PF  
 XX 17-DEC-1993; 93US-0168809.  
 PR  
 XX (PION-) PIONEER HI-BRED INT INC.  
 PA  
 XX Putman RJ, Rao AG, Rao A;  
 PI  
 XX WPI; 1995-231570/30.  
 DR  
 XX

XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 PT used against plant pathogenic fungi or human or animal infections  
 XX

PS Claim 1; Page 30; 45pp; English.

XX AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.  
 XX  
 XX Sequence 17 AA;

Query Match 78.0%; Score 71; DB 16; Length 17;  
 Best Local Similarity 70.6%; Pred. No. 3.5e-05;  
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
 : | ||| |||| ||| |  
 Db 1 kwlfvnyrgikyrrqr 17

RESULT 10  
 AAR75808  
 ID AAR75808 standard; peptide; 17 AA.  
 AC  
 AC AAR75808;  
 XX  
 XX 07-FEB-1996 (first entry)  
 DT

XX Antimicrobial tachyplesin peptide derivative.  
 DE  
 XX Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 KW infection.  
 KW  
 XX Synthetic.  
 OS  
 XX WO9516776-A1.  
 PN  
 XX 22-JUN-1995.  
 PD  
 XX 19-DEC-1994; 94WO-US14619.  
 PF  
 XX 17-DEC-1993; 93US-0168809.  
 PR  
 XX (PION-) PIONEER HI-BRED INT INC.  
 PA

XX Putman RJ, Rao AG, Rao A;  
 PI  
 XX WPI; 1995-231570/30.  
 DR  
 XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 PT used against plant pathogenic fungi or human or animal infections  
 XX  
 XX Claim 1; Page 30; 45pp; English.

XX AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.  
 XX  
 XX Sequence 17 AA;

Query Match 78.0%; Score 71; DB 16; Length 17;  
 Best Local Similarity 70.6%; Pred. No. 3.5e-05;

KW	Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;
KW	infection.
XX	Synthetic.
XX	WO9516776-A1.
XX	22-JUN-1995.
PD	XX
XX	19-DEC-1994; 94WO-US14619.
XX	17-DEC-1993; 93US-0168809.
XX	(PION-) PIONEER HI-BRED INT INC.
XX	Putman RJ, Rao AG, Rao A;
PPI	WPI; 1995-231570/30.
DR	XX
XX	New peptide derivs. of tachyplesin - having antimicrobial activity,
VPT	used against plant pathogenic fungi or human or animal infections
PPT	Claim 1; Page 31; 45pp; English.
XX	AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin
CC	(AAR75805) a small peptide isolated from Japanese horseshoe crab
CC	haemocytes which has antimicrobial properties. The peptide derivatives
CC	also have antimicrobial activity and can be used for killing and
CC	inhibiting fungi, in particular for fungi pathogenic to plants e.g.
CC	Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,
CC	Sclerotinia trifoliorum and Aspergillus flavus. The peptides can
CC	also be used for treating and preventing infection in humans and
CC	animals.
XX	Sequence 35 AA;
SQ	
Query Match 78.08; Score 71; DB 16; Length 35;	
Best Local Similarity 70.6%; Pred. No. 7.4e-05;	
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;	
QY	1 RMSFRVSVRGISYRRSR 17
DB	:       :             1 kwlfrvnyrgikyrrqr 17
RESULT 13	
AAR75822	
ID	AAR75822 standard; peptide: 17 AA.
XX	
AC	AAR75822;
XX	
DT	07-FEB-1996 (first entry)
XX	
DE	Antimicrobial tachyplesin peptide derivative.
XX	
KW	Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;
KW	infection.
XX	Synthetic.
OS	WO9516776-A1.
PN	22-JUN-1995.
XX	
PD	XX
XX	19-DEC-1994; 94WO-US14619.
PF	17-DEC-1993; 93US-0168809.
PR	(PION-) PIONEER HI-BRED INT INC.
XX	Putman RJ, Rao AG, Rao A;
PA	
XX	
PI	
XX	

DR WPI: 1995-231570/30.  
 XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 PT used against plant pathogenic fungi or human or animal infections  
 XX  
 PS Claim 1; Page 36; 45pp; English.  
 XX  
 CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.  
 XX  
 SQ Sequence 17 AA;  
 Query Match 76.9%; Score 70; DB 16; Length 17;  
 Best Local Similarity 70.6%; Pred. No. 5.1e-05;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 RWSFRVSYRGISYRRSR 17  
 : | | | | | | | | | |  
 Db 1 kwrfvryrgieyrrer 17  
 RESULT 14  
 AAP91671  
 ID AAP91671 standard; peptide; 17 AA.  
 AC AAP91671;  
 XX  
 DT 29-JUN-1990 (first entry)  
 XX  
 DE New lipopolysaccharide-binding polypeptide(s).  
 XX  
 KW Lipopolysaccharide (LPS) endotoxins; bacterial infections;  
 KW lipopolysaccharide (LPS) endotoxins; antibacterial agents;  
 KW LPS-mediated immune disorders; inflammatory disorders;  
 KW horseshoe crab haemocytes.  
 XX  
 OS Horseshoe crab.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1 /label=OTHER  
 FT /note="H-Lys"  
 FT Disulfide-bond 3..16  
 FT Disulfide-bond 7..12  
 FT Misc-difference 17 /label=OTHER  
 FT /note="Arg-OH or Arg-NH2"  
 XX  
 PN W08901492-A.  
 XX  
 PD 23-FEB-1989.  
 XX  
 PF 19-AUG-1988; 88WO-JP00823.  
 XX  
 PR 21-AUG-1987; 87JP-0206258.  
 XX  
 PA (SEK ) SEIKAGAKU KOGYO KK.  
 XX  
 XX Nakamura T, Iwanaga S, Ohno M, Miyazaki K;  
 PI WPI: 1989-068854/09.  
 XX  
 XX New lipo:polysaccharide- binding polypeptide(s) -  
 PT useful for treating bacterial infections and immune and  
 PT inflammatory disorders.

XX  
 PS Claim 2; Page 27; 39pp; English.  
 XX  
 CC The lipopolysaccharide-binding polypeptides may be prepd. by either  
 CC solid-phase peptide synthesis followed by oxidn. to form the disulphide  
 CC bridges or, subjecting horseshoe crab haemocytes to hypotonic extraction,  
 CC extracting the residue with acid, and purifying the extract. The  
 CC polypeptides have high affinity for lipopolysaccharide (LPS) endotoxins  
 CC and are useful for removing such toxins from fluids, as antibacterial  
 CC agents, eg active against Salmonella spp. and S. aureus, and for  
 CC treating LPS-mediated immune and inflammatory disorders, eg superior  
 CC tracheobronchial infections, urinary tract infections, bedsores, burns,  
 CC colitis, cirrhosis, hepatic insufficiency and post-operative  
 CC complications.  
 XX  
 SQ Sequence 17 AA;

Query Match 74.7%; Score 68; DB 10; Length 17;  
 Best Local Similarity 70.6%; Pred. No. 0.00011;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
 : | | | | | | | | | |  
 Db 1 kwcfrcvrgicyrrer 17

RESULT 15  
 AAR06266  
 ID AAR06266 standard; peptide; 17 AA.  
 AC AAR06266;  
 XX  
 DT 13-DEC-1990 (first entry)  
 XX  
 DE Antiviral peptide.  
 XX  
 DE Vesicular stomatitis virus; HIV; AIDS;  
 XX  
 OS Tachypaus tridentatus.  
 XX  
 FH Key Location/Qualifiers  
 FT Disulfide-bond 3..16  
 FT Disulfide-bond 7..12  
 XX  
 PN JP02167230-A.  
 XX  
 PD 27-JUN-1990.  
 XX  
 PF 30-JUN-1989; 89JP-0166811.  
 XX  
 PR 26-SEP-1988; 88JP-0239051.  
 PR 30-JUN-1989; 89JP-0166811.  
 XX  
 PA (SEK ) SEIKAGAKU KOGYO KK.  
 XX  
 XX WPI: 1990-241996/32.  
 XX  
 PT Antivirus agents of polypeptide - useful as antiviral agents for  
 PT vesicular stomatitis virus or human immuno-deficiency virus  
 XX  
 PS Claim 1; Page 309; 12pp; Japanese.  
 XX  
 SQ Sequence 17 AA;

Query Match 74.7%; Score 68; DB 11; Length 17;  
 Best Local Similarity 70.6%; Pred. No. 0.00011;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
 : | | | | | | | | | |  
 Db 1 kwcfrcvrgicyrrer 17

Search completed: February 12, 2002, 12:30:33  
Job time: 366 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:24 ; Search time 106.12 seconds  
(without alignments)  
3.605 Million cell updates/sec

Title: US-09-485-571-27  
Perfect score: 91  
Sequence: 1 RWSFRVSYRGISYRRSR 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	83.5	17	1	US-08-168-809-5
2	72	79.1	17	1	US-08-168-809-18
3	72	79.1	17	3	US-08-962-034-2
4	71	78.0	17	1	US-08-168-809-6
5	71	78.0	17	1	US-08-168-809-7
6	71	78.0	21	1	US-08-168-809-15
7	71	78.0	35	1	US-08-168-809-9
8	70	76.9	17	1	US-08-168-809-21
9	68	74.7	17	1	US-07-926-965-1
10	68	74.7	17	1	US-07-876-883-1
11	68	74.7	17	1	US-07-876-883-2
12	68	74.7	17	1	US-08-168-809-4
13	68	74.7	17	1	US-08-426-550-1
14	68	74.7	17	1	US-08-426-550-2
15	68	74.7	17	3	US-08-962-034-1
16	68	74.7	39	1	US-08-168-809-16
17	67	73.6	17	1	US-08-168-809-2
18	67	73.6	17	1	US-08-168-809-14
19	65	71.4	17	1	US-07-876-883-3
20	65	71.4	17	1	US-08-426-550-3
21	64	70.3	17	1	US-08-168-809-12
22	64	70.3	17	1	US-08-168-809-13
23	64	70.3	17	1	US-08-168-809-19
24	64	70.3	18	1	US-07-876-883-4
25	64	70.3	18	1	US-08-282-030-7
26	64	70.3	18	1	US-08-426-550-4
27	64	70.3	18	5	PCT-US95-10219-7

28	64	70.3	19	1	US-08-282-030-8	Sequence 8, Appli
29	64	70.3	19	5	PCT-US95-10219-8	Sequence 8, Appli
30	62	68.1	17	1	US-07-856-026B-13	Sequence 13, Appli
31	62	68.1	17	1	US-08-168-809-10	Sequence 10, Appli
32	62	68.1	18	1	US-07-856-026B-14	Sequence 14, Appli
33	61	67.0	17	1	US-08-168-809-11	Sequence 11, Appli
34	61	67.0	18	1	US-08-037-777A-1	Sequence 1, Appli
35	61	67.0	18	1	US-07-876-883-5	Sequence 5, Appli
36	61	67.0	18	1	US-07-856-026B-23	Sequence 23, Appli
37	61	67.0	18	1	US-08-379-039C-1	Sequence 1, Appli
38	61	67.0	18	1	US-08-426-550-5	Sequence 5, Appli
39	61	67.0	18	2	US-08-459-400-1	Sequence 1, Appli
40	60	65.9	17	4	US-09-230-180-36	Sequence 36, Appli
41	59	64.8	17	1	US-07-856-026B-3	Sequence 3, Appli
42	59	64.8	18	1	US-07-856-026B-4	Sequence 4, Appli
43	58	63.7	17	1	US-07-856-026B-21	Sequence 21, Appli
44	57	62.6	16	1	US-07-856-026B-12	Sequence 12, Appli
45	57	62.6	18	1	US-07-856-026B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-168-809-5  
; Sequence 5, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; APPLICANT: Rao, Aragula G.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
; INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Stree  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,809  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roth, Michael J.  
; REGISTRATION NUMBER: 29,342  
; REFERENCE/DOCKET NUMBER: 0173R US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 515-245-3595  
; TELEFAX: 515-245-3634  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-168-809-5

Query Match 83.5% Score 76; DB 1; Length 17;  
Best Local Similarity 70.6% Pred. No. 1.3e-06;  
Matches 12; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RWSFRVSYRGISYRRSR 17

```

QY      1 RWSFRVSYRGISYRRSR 17
      : 1 111 1111 1111 1
Db      1 KWXFRVXYRGIXYRRXR 17

      4
US-08-168-809-6
; Sequence 6, Application US/0816809
; Patent No. 5580852
; GENERAL INFORMATION:
; APPLICANT: Putnam, Rebecca J.
; APPLICANT: Rao, Aragula G.
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING
; INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pioneer Hi-Bred International
; STREET: 700 Capital Square, 400 Locust Stree
; CITY: Des Moines
; STATE: IA
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/168,809
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Roth, Michael J.
; REGISTRATION NUMBER: 29,342
; REFERENCE/DOCKET NUMBER: 0173R US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-245-3595
; TELEFAX: 515-245-3634
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid

```

2

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-168-809-6

Query Match 78.0%; Score 71; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 8.9e-06;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RNSFRVSYRGISYRRSR 17  
: | | | | | | | | | |  
Db 1 KWLFRVNYRGIKYRRQR 17

RESULT 5  
US-08-168-809-7  
; Sequence 7, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; APPLICANT: Rao, Aragula G.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,809  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roth, Michael J.  
; REGISTRATION NUMBER: 29,342  
; REFERENCE/DOCKET NUMBER: 0173R US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 515-245-3595  
; TELEFAX: 515-245-3634  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-168-809-7

Query Match 78.0%; Score 71; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 8.9e-06;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RNSFRVSYRGISYRRSR 17  
: | | | | | | | | | |  
Db 1 KWLFRVNYRGIKYRRQR 17

RESULT 6  
US-08-168-809-15

; Sequence 15, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; APPLICANT: Rao, Aragula G.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,809  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roth, Michael J.  
; REGISTRATION NUMBER: 29,342  
; REFERENCE/DOCKET NUMBER: 0173R US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 515-245-3595  
; TELEFAX: 515-245-3634  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-168-809-15

Query Match 78.0%; Score 71; DB 1; Length 21;  
Best Local Similarity 70.6%; Pred. No. 1.1e-05;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RNSFRVSYRGISYRRSR 17  
: | | | | | | | | | |  
Db 1 KWLFRVNYRGIKYRRQR 17

RESULT 7  
US-08-168-809-9  
; Sequence 9, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; APPLICANT: Rao, Aragula G.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Street  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

us-09-485-571-27.rai

Wed Feb 13 07:52:25 2002

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/168,809
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Roth, Michael J.
; REGISTRATION NUMBER: 29,342
; REFERENCE/DOCKET NUMBER: 0173R US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-245-3595
; TELEFAX: 515-245-3634
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
US-08-168-809-9

```

Query Match 78.0%; Score 71; DB 1; Length 35;  
 Best Local Similarity 70.6%; Pred. No. 1.9e-05; Indels 0;  
 Matches 12; Conservative 2; Mismatches 3; Gaps 0;

```

QY 1 RWSFRVSYRGISYRSR 17
DB 1 KWLFRVNYGKRYRQR 17

```

```

RESULT 8
US-08-168-809-21
; Sequence 21, Application US/08168809
; Patent No. 5580852
; GENERAL INFORMATION:
; APPLICANT: Putnam, Rebecca J.
; APPLICANT: Rao, Aragula G.
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLEISIN' HAVING
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pioneer Hi-Bred International
; STREET: 700 Capital Square, 400 Locust Stree
; CITY: Des Moines
; STATE: IA
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/168,809
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Roth, Michael J.
; REGISTRATION NUMBER: 29,342
; REFERENCE/DOCKET NUMBER: 0173R US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-245-3595
; TELEFAX: 515-245-3634
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
US-08-168-809-21
;
; Query Match 76.9%; Score 70; DB 1; Length 17;
; Best Local Similarity 70.6%; Pred. No. 1.3e-05; Indels 0;
; Matches 12; Conservative 1; Mismatches 4; Gaps 0;
;
QY 1 RWSFRVSYRGISYRSR 17
DB 1 KWRFRVRYRGIEYRER 17

```

```

RESULT 9
US-07-926-965-1
; Sequence 1, Application US/07926965
; Patent No. 5416194
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, TAKANORI; IWANAGA, SADAOKI;
; APPLICANT: OHNO, MOTONORI; MIYAZAKI, KYOSUKE
; TITLE OF INVENTION: NOVEL POLYPEPTIDE AND METHOD FOR
; TITLE OF INVENTION: PREPARING THE SAME
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/926,965
; FILING DATE: 19920807
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/665,819
; FILING DATE: 07-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/348,487
; FILING DATE: 19-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: TSU-4B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-661-8000
; TELEFAX: 212-661-8002
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY:
; MOLECULE TYPE: POLYPEPTIDE
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HORSESHOE CRAB
; STRAIN: TACHYPLEUS TRIDENTATUS
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE: HEMOCYTE
; CELL LINE:
; ORGANELLE:
; FEATURE:
; NAME/KEY: LPS-binding polypeptide, or LPB

```

;  
; LOCATION:  
; IDENTIFICATION METHOD:  
; OTHER INFORMATION: WHEREIN CYS-3 AND CYS-16 CAN  
; OTHER INFORMATION: FORM A DISULFIDE BOND, AND CYS-7 AND CYS-12 CAN FORM A DISULF  
; OTHER INFORMATION: BOND; WHEREIN CARBOXYL GROUP OF ARG-17 CAN BE AMIDATED; WHERE  
; OTHER INFORMATION: OR ALL OF THE RESIDUES CAN BE PROTECTED WITH PROTECTIVE GROU  
US-07-926-965-1

Query Match 74.7%; Score 68; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.8e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RWSFRVSYRGISYRRSR 17  
: | | | | | | | | | |  
Db 1 KWCPRVCYRGICYRRCR 17

RESULT 10  
US-07-876-883-1  
; Sequence 1, Application US/07876883  
; Patent No. 5449752  
; GENERAL INFORMATION:  
; APPLICANT: Fujii, No. 5449752utaka  
; APPLICANT: Yamamoto, Naoki  
; APPLICANT: Matsumoto, Akiyoshi  
; APPLICANT: Waki, Michinori  
; TITLE OF INVENTION: No. 5449752el Polypeptides With Affinity To  
; TITLE OF INVENTION: Lipopolysaccharides And Their Uses  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/876,883  
; FILING DATE: 19920429  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Miller, Charles E.  
; REGISTRATION NUMBER: 24,576  
; REFERENCE/DOCKET NUMBER: 7568-003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-07-876-883-1.

Query Match 74.7%; Score 68; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.8e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RWSFRVSYRGISYRRSR 17  
: | | | | | | | | | |  
Db 1 KWCPRVCYRGICYRRCR 17

RESULT 11  
US-07-876-883-2  
; Sequence 2, Application US/07876883  
; Patent No. 5449752  
; GENERAL INFORMATION:  
; APPLICANT: Fujii, No. 5449752utaka  
; APPLICANT: Yamamoto, Naoki  
; APPLICANT: Matsumoto, Akiyoshi  
; APPLICANT: Waki, Michinori  
; TITLE OF INVENTION: No. 5449752el Polypeptides With Affinity To  
; TITLE OF INVENTION: Lipopolysaccharides And Their Uses  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/876,883  
; FILING DATE: 19920429  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Miller, Charles E.  
; REGISTRATION NUMBER: 24,576  
; REFERENCE/DOCKET NUMBER: 7568-003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-07-876-883-2

Query Match 74.7%; Score 68; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.8e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RWSFRVSYRGISYRRSR 17  
: | | | | | | | | | |  
Db 1 KWCPRVCYRGICYRRCR 17

RESULT 12  
US-08-168-809-4  
; Sequence 4, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; APPLICANT: Rao, Aragula G.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Stree  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/168,809  
;; FILING DATE:  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Roth, Michael J.  
;; REGISTRATION NUMBER: 29,342  
;; REFERENCE/DOCKET NUMBER: 0173R US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 515-245-3595  
;; TELEFAX: 515-245-3634  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 17 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
US-08-168-809-4

Query Match 74.7%; Score 68; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.8e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
Db 1 KWCPRVCYRGICYYRCR 17

RESULT 13  
US-08-426-550-1  
; Sequence 1, Application US/08426550  
; Patent No. 5710128  
; GENERAL INFORMATION:  
; APPLICANT: Fujii, No. 5710128utaka  
; APPLICANT: Yamamoto, Naoki  
; APPLICANT: Matsumoto, Akiyoshi  
; APPLICANT: Waki, Michinori  
; TITLE OF INVENTION: Pharmaceutical Compositions of  
; LIPOLYSACCHARIDE-BINDING POLYPEPTIDES  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,550  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Miller, Charles E.  
; REGISTRATION NUMBER: 24,576  
; REFERENCE/DOCKET NUMBER: 7568-006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 869-9741  
; TELEFAX: 212 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-426-550-1

Query Match 74.7%; Score 68; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.8e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
Db 1 KWCPRVCYRGICYYRCR 17

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 17 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: peptide  
US-08-426-550-1

Query Match 74.7%; Score 68; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.8e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
Db 1 KWCPRVCYRGICYYRCR 17

RESULT 14  
US-08-426-550-2  
; Sequence 2, Application US/08426550  
; Patent No. 5710128  
; GENERAL INFORMATION:  
; APPLICANT: Fujii, No. 5710128utaka  
; APPLICANT: Yamamoto, Naoki  
; APPLICANT: Matsumoto, Akiyoshi  
; APPLICANT: Waki, Michinori  
; TITLE OF INVENTION: Pharmaceutical Compositions of  
; LIPOLYSACCHARIDE-BINDING POLYPEPTIDES  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,550  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Miller, Charles E.  
; REGISTRATION NUMBER: 24,576  
; REFERENCE/DOCKET NUMBER: 7568-006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 869-9741  
; TELEFAX: 212 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-426-550-2

Query Match 74.7%; Score 68; DB 1; Length 17;  
Best Local Similarity 70.6%; Pred. No. 2.8e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
Db 1 KWCPRVCYRGICYYRCR 17

```

RESULT 15
US-08-962-034-1
; Sequence 1, Application US/08962034
; Patent No. 6015941
; GENERAL INFORMATION:
; APPLICANT: Rao, A. Gururaj
; TITLE OF INVENTION: PEPTIDE DERIVATIVES OF TACHYPLESIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer IP Group of Alston & Bird
; STREET: 3605 Glenwood Ave.
; CITY: Raleigh
; STATE: NC
; COUNTRY: US
; ZIP: 27622-1107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/962,034
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5718-19
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-962-034-1

```

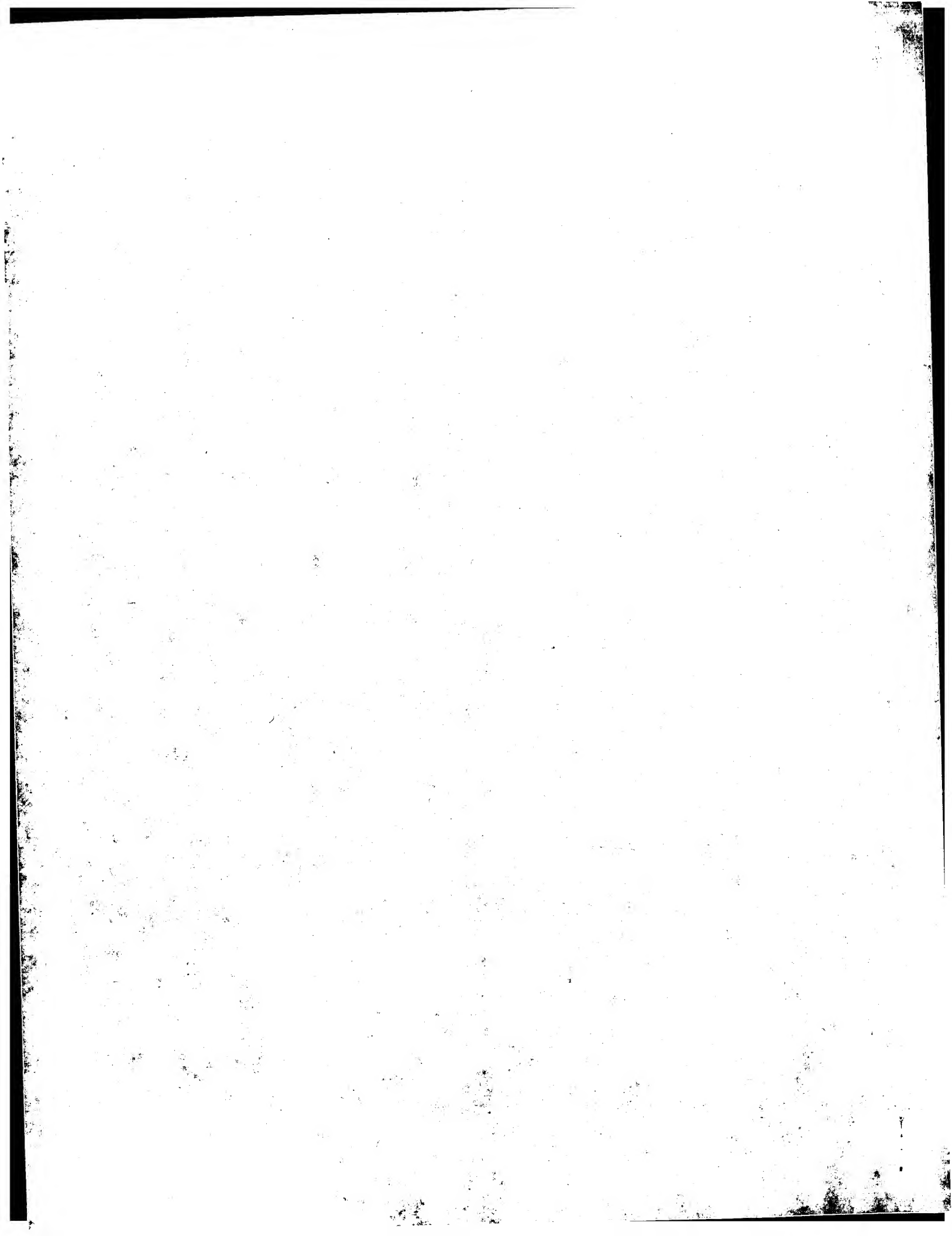
```

Query Match      74.7%; Score 68; DB 3; Length 17;
Best Local Similarity 70.6%; Pred. No. 2.8e-05;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17
   :| ||| ||| ||| |
DB 1 KWCFRVCYRGICYRRCR 17

```

Search completed: February 12, 2002, 12:32:24  
Job time: 452 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:41 ; Search time 126.85 Seconds  
(without alignments)  
10.209 Million cell updates/sec

Title: US-09-485-571-27  
Perfect score: 91  
Sequence: 1 RWSFRVSYRGISYRRSR 17  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	74.7	17	A38824	tachyplesin I - ho
2	68	74.7	19	JX0124	tachyplesin I prec
3	68	74.7	77	A38345	tachyplesin I prec
4	68	74.7	77	B38345	tachyplesin II pre
5	65	71.4	17	JX0125	tachyplesin III -
6	64	70.3	18	JU0124	polypeptidase II - A
7	61	67.0	18	JU0125	polypeptidase II -
8	47	51.6	615	D86473	hypothetical prote
9	44.5	48.9	88	C82472	conserved hypothet
10	44	48.4	620	E86468	protein F1X21.26
11	42	46.2	480	G75313	molybdate metabol
12	42	46.2	536	D42463	hypothetical prote
13	41	45.1	94	D45928	MHC HLA-DQ-beta ce
14	41	45.1	261	HLH92C	MHC class II histo
15	41	45.1	279	T02495	hypothetical prote
16	41	45.1	307	F84162	hypothetical prote
17	41	45.1	378	A84161	hypothetical prote
18	41	45.1	720	A36942	Fe(III)-pyochelin
19	41	45.1	770	T22808	hypothetical prote
20	40	44.0	90	I59639	MHC class II histo
21	40	44.0	108	I72482	HLA DR-beta-I. hu
22	40	44.0	261	I68718	MHC class II histo
23	40	44.0	330	A72534	hypothetical prote
24	40	44.0	378	G84093	integrase (phage-r
25	40	44.0	732	S25995	hypothetical prote
26	40	44.0	901	H82850	outer membrane ush
27	40	44.0	1099	S48053	protein tyrosine k
28	40	44.0	1100	S43677	protein tyrosine k
29	40	44.0	1124	A55747	L-JAK protein-tyro

30 40 44.0 1299 2 I58401 protein-tyrosine k  
31 39.5 43.4 781 2 A86205 hypothetical prote  
32 39 42.9 242 2 D72485 probable high-affi  
33 39 42.9 413 2 T39168 probable agmatinas  
34 39 42.9 767 2 G86476 protein F1504.37 l  
35 39 42.9 968 2 S48992 protein p130 - rat  
36 39 42.9 1148 2 F86403 probable transposo  
37 38 41.8 203 2 C25511 cc protein - fruit  
38 38 41.8 225 2 S68196 hypothetical prote  
39 38 41.8 329 1 D71316 conserved hypothet  
40 38 41.8 330 2 E96503 protein F9C16.11 l  
41 38 41.8 358 2 H75264 hypothetical prote  
42 38 41.8 411 2 S46800 LAG1 protein - yea  
43 38 41.8 466 2 E75201 pyridoxal phosphat  
44 38 41.8 514 2 B72752 hypothetical prote  
45 38 41.8 787 2 A55034 6-phosphofructokin

ALIGNMENTS

RESULT 1  
A38824  
tachyplesin I - horseshoe crab (Tachyplesus gigas)  
C:Species: Tachyplesus gigas  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997  
C:Accession: A38824  
R:Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.  
J. Biochem. 108, 261-266, 1990  
A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (C  
ssing intermediate of its precursor.  
A:Reference number: JX0124; MUID:91035357  
A:Accession: A38824  
A:Molecule type: protein  
A:Residues: 1-17 <MUT>  
A:Experimental source: hemocyte  
C:Keywords: amidated carboxyl end  
F:3-16,7-12/Disulfide bonds: #status predicted  
F:17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 74.7%; Score 68; DB 2; Length 17;  
Best Local Similarity 70.6%; Pred. No. 5.5e-05;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
: | | | | | | | | | | | | | | | |  
Db 1 KWCFRVCYRGICVRRCR 17

RESULT 2  
JX0124  
tachyplesin I precursor - horseshoe crab (Carinoscorpius rotundicauda)  
C:Species: Carinoscorpius rotundicauda  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997  
C:Accession: JX0124  
R:Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.  
J. Biochem. 108, 261-266, 1990  
A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (C  
ssing intermediate of its precursor.  
A:Reference number: JX0124; MUID:91035357  
A:Accession: JX0124  
A:Molecule type: protein  
A:Residues: 1-19 <MUT>  
A:Experimental source: hemocyte  
C:Keywords: amidated carboxyl end  
F:1-17/precursor: tachyplesin I #status experimental <MAT>  
F:3-16,7-12/Disulfide bonds: #status predicted  
F:17/Modified site: amidated carboxyl end (Arg) (amide in mature form from followin

Query Match 74.7%; Score 68; DB 2; Length 19;  
Best Local Similarity 70.6%; Pred. No. 6.1e-05;

Wed Feb 13 07:52:26 2002

us-09-485-571-27.rpr

Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
: || || | || || || || |  
DB 1 KWCFRVCYRGICYRKCR 17

RESULT 3

A38345  
Tachyplesin I precursor - horseshoe crab (Tachyplesus tridentatus)  
C:Species: Tachyplesus tridentatus  
C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 01-Dec-2000  
C:Accession: A38345; A30068  
R:Shigenaga, T.; Muta, T.; Toh, Y.; Tokunaga, F.; Iwanaga, S.  
J. Biol. Chem. 265, 21350-21354, 1990  
A:Title: Antimicrobial tachyplesin peptide precursor. cDNA cloning and cellular localization  
A:Reference number: A38345; MUID:91065956  
A:Accession: A38345  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-77 <SHI>  
A:Cross-references: GB:M57242; GB:J05689; NID:gl161659; PID:gl161660  
R:Nakamura, T.; Furunaka, H.; Miyata, T.; Tokunaga, F.; Muta, T.; Iwanaga, S.; Niwa, M.; Takao, T.;  
J. Biol. Chem. 263, 16709-16713, 1988  
A:Title: Tachyplesin, a class of antimicrobial peptide from the hemocytes of the horseshoe crab  
A:Reference number: A30068; MUID:89034158  
A:Accession: A30068  
A:Molecule type: protein  
A:Residues: 24-40 <NAK>

Query Match 74.7%; Score 68; DB 2; Length 77;

Best Local Similarity 70.6%; Pred. No. 0.00025;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
: || || | || || || || |  
DB 24 KWCFRVCYRGICYRKCR 40

RESULT 4

B38345  
Tachyplesin II precursor - horseshoe crab (Tachyplesus tridentatus)  
C:Species: Tachyplesus tridentatus  
C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 08-Dec-2000  
C:Accession: B38345; JU0123  
R:Shigenaga, T.; Muta, T.; Toh, Y.; Tokunaga, F.; Iwanaga, S.  
J. Biol. Chem. 265, 21350-21354, 1990  
A:Title: Antimicrobial tachyplesin peptide precursor. cDNA cloning and cellular localization  
A:Reference number: A38345; MUID:91065956  
A:Accession: B38345  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-77 <SHI>  
A:Cross-references: GB:J05689  
R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, T.;  
J. Biochem. 106, 663-668, 1989  
A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin II,  
A:Reference number: A91914; MUID:90110066  
A:Accession: JU0123  
A:Molecule type: protein  
A:Residues: 24-40 <MIY>  
A:Comment: The peptide is one of the antimicrobial peptides found in the Japanese horseshoe crab hemocytes, isolated from horseshoe crab hemocytes, tachyplesin II,  
C:Keywords: amidated carboxyl end  
F:26-39,30-35/Disulfide bonds: #status predicted  
F:40/Modified site: amidated carboxyl end (Arg) (amide in mature form from following gly

Query Match 74.7%; Score 68; DB 2; Length 77;  
Best Local Similarity 70.6%; Pred. No. 0.00025;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17

DB 24 RWCFRVCYRGICYRKCR 40

RESULT 5

JX0125  
Tachyplesin III - horseshoe crab (Tachyplesus gigas)  
C:Species: Tachyplesus gigas  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997  
C:Accession: JX0125  
R:Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.  
J. Biochem. 108, 261-266, 1990  
A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (Ca  
ssing intermediate of its precursor  
A:Reference number: JX0124; MUID:91035357  
A:Accession: JX0125  
A:Molecule type: protein  
A:Residues: 1-17 <MUT>  
A:Experimental source: hemocyte  
A:Keywords: amidated carboxyl end  
F:3-16,7-12/Disulfide bonds: #status predicted  
F:17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 71.4%; Score 65; DB 2; Length 17;

Best Local Similarity 64.7%; Pred. No. 0.00017;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
: || || | || || || || |  
DB 1 KWCFRVCYRGICYRKCR 17

RESULT 6

JU0124  
polypheus I - Atlantic horseshoe crab  
C:Species: Limulus polyphemus (Atlantic horseshoe crab)  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jul-1997  
C:Accession: JU0124  
R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J.  
J. Biochem. 106, 663-668, 1989  
A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin  
A:Reference number: A91914; MUID:90110066  
A:Accession: JU0124  
A:Molecule type: protein  
A:Residues: 1-18 <MIY>  
A:Comment: The peptide is one of the antimicrobial peptides in the Atlantic horseshoe  
C:Keywords: amidated carboxyl end  
F:4-17,8-13/Disulfide bonds: #status experimental  
F:18/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 70.3%; Score 64; DB 2; Length 18;

Best Local Similarity 64.7%; Pred. No. 0.00026;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRRSR 17  
: || || | || || || || |  
DB 2 RWCFRVCYRGICYRKCR 18

RESULT 7

JU0125  
polypheus II - Atlantic horseshoe crab  
C:Species: Limulus polyphemus (Atlantic horseshoe crab)  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jul-1997  
C:Accession: JU0125  
R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J.  
J. Biochem. 106, 663-668, 1989  
A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin  
A:Reference number: A91914; MUID:90110066  
A:Accession: JU0125  
A:Molecule type: protein

; Al.  
ewar  
; Kl  
Marz

Tal

n R1

n, R  
1, C

S R1

**D: 96**

Wed Feb 13 07:52:26 2002

QY 2 WSRVSYRGISYRRSR 17  
| | | | | | | | | |  
Db 234 WMRVDARGVYGHSR 249  
| | | | | | | | | |

RESULT 12  
D42463  
hypothetical protein Bcv' (pinB 5' region) - Shigella boydii (fragment)  
C:Species: Shigella boydii  
C:Date: 10-Jul-1992 #sequence\_revision 18-Sep-1992 #text\_change 30-Sep-1993  
C:Accession: D42463  
R:Tomlinaga, A.; Ikemizu, S.; Enomoto, M.  
J. Bacteriol. 173, 4079-4087, 1991  
A:Title: Site-specific recombinase genes in three Shigella subgroups and nucleotide sequence  
A:Reference number: A42463; MUID:91286192  
A:Accession: D42463  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-536 <TOM>  
A:Cross-references: GB:D00660

Query Match 46.2%; Score 42; DB 2; Length 536;  
Best Local Similarity 62.5%; Pred. No. 32;  
Matches 10; Conservative 2; Mismatches 2; Gaps 1;  
Indels 2

QY 4 FRVSYR--GISYRRSR 17  
| | | | | | | | | |  
Db 423 FRVNRGGIFYSR 438  
| | | | | | | | | |

RESULT 13  
159528  
MHC HLA-DQ-beta cell surface glycoprotein - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 23-Jul-1999  
C:Accession: 159528  
R:Sinha, A.A.; Brautbar, C.; Szafer, F.; Friedmann, A.; Tzfoni, E.; Todd, J.A.; Steinman  
Science 239, 1026-1029, 1988  
A:Title: A newly characterized HLA DQ beta allele associated with pemphigus vulgaris.  
A:Reference number: 159528; MUID:98145646  
A:Accession: 159528  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-94 <RES>  
A:Cross-references: GB:M19239; NID:g181746; PIDN:AAA52319.1; PID:g181747  
C:Genetics:  
A:Gene: GDB:HLA-DQB1  
A:Cross-references: GDB:120517; OMIM:142857  
A:Map position: 6p21.3-6p21.3  
A:Superfamily: class II histocompatibility antigen; immunoglobulin homology  
C:Keywords: glycoprotein

Query Match 45.1%; Score 41; DB 2; Length 94;  
Best Local Similarity 53.3%; Pred. No. 8.2;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRR 15  
| | | | | | | | | |  
Db 80 RHNVEVAYRGILQRR 94  
| | | | | | | | | |

RESULT 14  
HLHU2C  
MHC class II histocompatibility antigen HLA-DQ beta chain (DQ5) precursor - human  
N:Alternate names: HLA-DC beta  
C:Species: Homo sapiens (man)  
C:Date: 18-Apr-1994 #sequence\_revision 02-May-1994 #text\_change 22-Jun-1999  
C:Accession: C24669; A02232; A60773; D33287; F33287; F33287; A35054; B35054; F35  
R:Tonnelle, C.; Demars, R.; Long, E.O.  
EMBO J. 4, 2839-2847, 1985  
A:Title: DO beta: a new beta chain gene in HLA-D with a distinct regulation of expressed

A:Reference number: A91020; MUID:86055719  
A:Accession: C24669  
A:Molecule type: mRNA  
A:Residues: 1-261 <TON>  
A:Cross-references: GB:X03058; NID:g32279; PIDN:CAA26872.1; PID:g32280  
A:Experimental source: clone DQw1.1  
R:Larhammar, D.; Andersson, G.; Andersson, M.; Bill, P.; Bohme, J.; Claesson, L.; De  
rvenius, B.; Widmark, E.; Rask, L.; Peterson, P.A.  
Hum. Immunol. 8, 95-103, 1983  
A:Title: Molecular analysis of human class II transplantation antigens and their ge  
A:Reference number: A91743; MUID:84031733  
A:Accession: A02232  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 33-261 <LAR>  
A:Experimental source: clone pII-beta-2  
R:Merlyman, P.; Silver, J.; Gregersen, P.K.; Solomon, G.; Winchester, R.  
J. Immunol. 143, 2068-2073, 1989  
A:Title: A novel association of DQalpha and DQbeta genes in the DRw10 haplotype. Det  
A:Reference number: A60773; MUID:89381351  
A:Accession: A60773  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 33-126 <MER>  
A:Experimental source: clone from DRw10 haplotype  
R:Scharf, S.J.; Freidmann, A.; Steinman, L.; Brautbar, C.; Erlich, H.A.  
Proc. Natl. Acad. Sci. U.S.A. 86, 6215-6219, 1989  
A:Title: Specific HLA-DQB and HLA-DRB1 alleles confer susceptibility to pemphigus v  
A:Reference number: A33287; MUID:89345634  
A:Accession: D33287  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 38-126 <SCJ>  
A:Experimental source: HLA-DQw1 beta-1.1, clone from DRw6, DQw1 haplotype  
A:Accession: E33287  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 53-88, 'S', '90-109 <SC2>  
A:Experimental source: allele designated DQB 1.2  
A:Accession: F33287  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 53-88, 'D', '90-109 <SC3>  
A:Experimental source: allele designated DQB 1.3  
A:Note: this allele appears to confer susceptibility to pemphigus vulgaris  
R:Gyllenstein, U.B.; Lashkari, D.; Erlich, H.A.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1835-1839, 1990  
A:Title: Allelic diversification at the class II DQB locus of the mammalian major h  
A:Reference number: A35054; MUID:90175391  
A:Accession: A35054  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 53-109 <GY2>  
A:Experimental source: allele designated DQB 1.1  
A:Accession: B35054  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 53-88, 'S', '90-109 <GY3>  
A:Experimental source: allele designated DQB 1.2  
A:Accession: C35054  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 53-88, 'D', '90-109 <GYL>  
A:Experimental source: allele designated DQB 1.3  
R:Horn, G.T.; Bugawan, T.L.; Long, C.M.; Erlich, H.A.  
Proc. Natl. Acad. Sci. U.S.A. 85, 6012-6016, 1988  
A:Title: Allelic sequence variation of the HLA-DQ loci: relationship to serology au  
A:Reference number: A35589; MUID:88320372  
A:Accession: F35589  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 53-88, 'D', '90-109 <HOR>  
A:Experimental source: allele designated DQB1\*05032



us-09-485-571-27.rpr

Wed Feb 13 07:52:26 2002

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:53 ; Search time 67.2 Seconds  
(without alignments)  
9.275 Million cell updates/sec

Title: US-09-485-571-27

Perfect score: 91  
Sequence: 1 RWSFRVSYRGISYRRSR 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	% Score	Length	ID	Description
1	68	74.7	17	TAC1_TACGI	P23684 tachyples
2	68	74.7	77	TAC1_TACTR	P14213 tachyples
3	68	74.7	77	TAC2_TACTR	P14214 tachyples
4	65	71.4	17	TAC3_TACGI	P18252 tachyples
5	64	70.3	18	PPM1_LIMPO	P14215 limulus pol
6	61	67.0	18	PPM2_LIMPO	P14216 limulus pol
7	41	45.1	261	HB22_HUMAN	P01919 homo sapien
8	41	45.1	720	FPTA_PSEAE	P42512 pseudomonas
9	40	44.0	732	YMI1_MARPO	P38456 marchantia
10	40	44.0	1100	JAK3_RAT	P63272 rattus norv
11	40	44.0	1124	JAK3_HUMAN	P52333 homo sapien
12	40	44.0	1299	JAK3_MOUSE	O62137 mus musculu
13	39	42.9	271	URED_ACTNA	Q92364 actinomycet
14	39	42.9	413	SPEC_SCHPO	O42887 schizosacch
15	39	42.9	968	BCA1_RAT	P63767 rattus norv
16	38	41.8	203	L2CC_DROME	P24156 drosophila
17	38	41.8	411	LAG1_YEAST	P38703 saccharomyc
18	38	41.8	616	ENP4_HUMAN	O9y227 homo sapien
19	38	41.8	787	K6PF_DROME	P52034 drosophila
20	38	41.8	1075	PLD1_RAT	P70496 rattus norv
21	37.5	41.2	288	VP30_EBOZM	O05323 ebola virus
22	37	40.7	120	RS13_BACSU	P20282 bacillus su
23	37	40.7	261	HB25_HUMAN	P03992 homo sapien
24	37	40.7	308	MACS_MOUSE	P26645 mus musculu
25	37	40.7	308	MACS_RAT	P30069 rattus norv
26	37	40.7	320	PR73_MMTVB	P10260 mouse mamma
27	37	40.7	473	DNB2_ADE41	P11806 human adeno
28	37	40.7	474	DNB2_ADE41	P11807 human adeno
29	37	40.7	529	Y632_CHLTR	O84637 chlamydia t
30	37	40.7	529	Y921_CHLMU	O9pb1 chlamydia m
31	37	40.7	721	MX_ANAPL	P33238 anas platyr
32	37	40.7	732	ACET_HUMAN	P22966 homo sapien
33	37	40.7	732	ACET_MOUSE	P22967 mus musculu

34	37	40.7	737	1	ACET_RABIT	P22968 oryctolagus
35	37	40.7	1306	1	ACE_HUMAN	P12821 homo sapien
36	37	40.7	1310	1	ACE_RABIT	P12822 oryctolagus
37	37	40.7	1312	1	ACE_MOUSE	P09470 mus musculu
38	37	40.7	1313	1	ACE_RAT	P47820 rattus norv
39	37	40.7	1453	1	NKCR_MOUSE	P30415 mus musculu
40	37	40.7	1462	1	NKCR_HUMAN	P30414 homo sapien
41	36	39.6	65	1	RL30_MYCTU	P95070 mycobacteri
42	36	39.6	71	1	RL30_MYCLE	O33001 mycobacteri
43	36	39.6	134	1	RL16_PINTH	P52767 pinus thunb
44	36	39.6	135	1	RK16_SPIOL	P17353 spinacia ol
45	36	39.6	175	1	KDGL_SYNY3	Q55143 synechocyst

ALIGNMENTS

RESULT 1	TAC1_TACGI	STANDARD;	PRT;	17 AA.
ID	TAC1_TACGI			
AC	P23684;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	TACHYPLESIN I.			
OS	Tachyplesus gigas (Southeast Asian horseshoe crab), and			
OS	Carcinoscorpius rotundicauda (Southeast Asian horseshoe crab).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;			
OC	Limulidae; Tachyplesus.			
OX	NCBI_TaxID=6852; 6848;			
RP	[1]			
RP	SEQUENCE.			
RC	SPECIES=T.gigas, and C.rotundicauda;			
RX	MEDLINE=91035357; PubMed=2225025;			
RA	Muta I.; Fujimoto T., Nakajima H., Iwanaga S.;			
RT	"Tachyplesus isolated from hemocytes of Southeast Asian horseshoe crabs (Carcinoscorpius rotundicauda and Tachyplesus gigas);			
RT	identification of a new tachyplesin, tachyplesin III, and a			
RT	processing intermediate of its precursor."			
RL	J. Biochem. 108:261-266(1990).			
CC	-I- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.			
DR	PIR; A38824; A38824.			
DR	PIR; JX0124; JX0124.			
KW	Antibiotic; Amidation.			
FT	DISULFID 3 16			
FT	DISULFID 7 12			
FT	MOD_RES 17 17			
SQ	SEQUENCE 17 AA; 2269 MW; E9E09BD9D923C94 CRC64;			
Query Match	74.7%;	Score 68;	DB 1;	Length 17;
Best Local Similarity	70.6%;	Pred. No. 1.1e-05;		
Matches 12;	Conservative 1;	Mismatches 4;	Indels 0;	Gaps 0;
Qy 1	RWSFRVSYRGISYRRSR 17			
Db 1	KWCFRCVYRGICVRRCR 17			
	:			
	:			
RESULT 2	TAC1_TACTR	STANDARD;	PRT;	77 AA.
ID	TAC1_TACTR			
AC	P14213;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	TACHYPLESIN I PRECURSOR.			
OS	Tachyplesus tridentatus (Japanese horseshoe crab).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;			
OC	Limulidae; Tachyplesus.			
OX	NCBI_TaxID=6853;			
RP	[1]			
RP	SEQUENCE FROM N.A.			

Wed Feb 13 07:52:27 2002

Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

RX MEDLINE=91065956; PubMed=2250028;  
 RA Shigenaga T., Muta T., Toh Y., Tokunaga F., Iwanaga S.;  
 RT "Antimicrobial tachyplesin peptide precursor. cDNA cloning and  
 RT cellular localization in the horseshoe crab (*Tachyplesus*  
 RT tridentatus).";  
 RL J. Biol. Chem. 265:21350-21354(1990).  
 RN [2]  
 RN SEQUENCE OF 24-40, AND DISULFIDE BONDS.  
 RP MEDLINE=89034158; PubMed=3141410;  
 RX Nakamura T., Furunaka H., Miyata T., Tokunaga F., Iwanaga S.,  
 RA Niwa M., Takao T., Shimonishi Y.;  
 RT "Tachyplesin, a class of antimicrobial peptide from the hemocytes of  
 RT the horseshoe crab (*Tachyplesus tridentatus*). Isolation and chemical  
 RT structure.";  
 RL J. Biol. Chem. 263:16709-16713(1988).  
 RN [3]  
 RN STRUCTURE BY NMR OF 24-40.  
 RP MEDLINE=90368729; PubMed=2394727;  
 RX Kawano K., Yonega T., Miyata T., Yoshikawa K., Tokunaga F.,  
 RA Terada Y., Iwanaga S.;  
 RT "Antimicrobial peptide, tachyplesin I, isolated from hemocytes of the  
 RT horseshoe crab (*Tachyplesus tridentatus*). NMR determination of the  
 RT beta-sheet structure.";  
 RL J. Biol. Chem. 265:15365-15367(1990).  
 RN [4]  
 RN STRUCTURE BY NMR OF 24-40.  
 RP MEDLINE=93257488; PubMed=8490053;  
 RX Tamamura H., Kuroda M., Masuda M., Otake A., Funakoshi S.,  
 RA Nakashima H., Yamamoto N., Waki M., Matsumoto A., Lancelin J.-M.,  
 RA Kohda D., Tate S., Inagaki F., Fujii N.;  
 RT "A comparative study of the solution structures of tachyplesin I and  
 RT a novel anti-HIV synthetic peptide, T22 ([Tyr5,12, Lys7]-polyphemusin  
 RT I), determined by nuclear magnetic resonance.";  
 RL Biochim. Biophys. Acta 1163:209-216(1993).  
 RN [5]  
 RN CHARACTERIZATION.  
 RP MEDLINE=94110249; PubMed=8282718;  
 RX Shigenaga T., Takayenoki Y., Kawasaki S., Seki N., Muta T., Toh Y.,  
 RA Ito A., Iwanaga S.;  
 RT "Separation of large and small granules from horseshoe crab  
 RT (tachyplesus tridentatus) hemocytes and characterization of their  
 RT components.";  
 RL J. Biochem. 114:307-316(1993).  
 CC -!- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND  
 CC GRAM-POSITIVE BACTERIA.  
 CC -!- SUBCELLULAR LOCATION: S-GRANULES.  
 CC -!- TISSUE SPECIFICITY: HEMOCYTES.  
 CC -!- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M57242; AAC63538.1;  
 CC PIR: A30068; A30068.  
 CC PIR: A38345; A38345.  
 CC Antibiatic; Amidation; Cleavage on pair of basic residues; Signal.  
 KW SIGNAL 1 23 TACHYPLESIN I.  
 FT PEPTIDE 24 40  
 FT PROPEP 41 77  
 FT DISULFID 26 39  
 FT DISULFID 30 35  
 FT MOD\_RES 40 40 AMIDATION (G-41 PROVIDE AMIDE GROUP).  
 FT DOMAIN 69 77 ASP/GLU-RICH (ACIDIC).  
 FT SEQUENCE 77 AA; 9349 MW; B940CAAA4641335F CRC64;  
 SQ  
 Query Match 74.7%; Score 68; DB 1; Length 77;  
 Best Local Similarity 70.6%; Pred. No. 5.le-05;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

..





39

```

ID YM11_MARPO STANDARD; PRT; 732 AA.
AC P38456;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 83.1 KDA PROTEIN IN COB-ATPA INTERGENIC REGION (ORF 732).
GN YMF11.
OS Marchantia polymorpha (Liverwort).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiales; Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114051; PubMed=1731062;
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyanaka K.;
RT "Gene organization deduced from the complete sequence of liverwort
RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant
RT mitochondrial genome.";
RL J. Mol. Biol. 223:1-7(1992).
CC -1- SIMILARITY: TO GROUP II INTRON MATURASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M68929; AAC09442.1; -
CC PIR: S25995; S25995.
CC Mendel: 2082; MARPO: ymf11.1.
CC InterPro: IPR000442; Intron_maturase2.
CC InterPro: IPR000477; RVSE.
CC Pfam: PF01348; Intron_maturase2; 1.
CC Pfam: PF00078; rvt; 1.
CC Mitochondrion; Hypothetical protein.
KW Mitochondrion; Hypothetical protein.
SQ SEQUENCE 732 AA; 83092 MW; 936CF036E9D06442 CRC64;

Query Match 44.08; Score 40; DB 1; Length 732;
Best Local Similarity 47.18; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 RWSFRVSRYGSIYRRSR 17
| | | | | | | |
Db 52 RLDFRMGGRAFSYERQR 68

RESULT 10
JAK3_RAT
ID JAK3_RAT STANDARD; PRT; 1100 AA.
AC Q63272;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TYROSINE-PROTEIN KINASE JAK3 (EC 2.7.1.112) (JANUS KINASE 3) (JAK-3).
GN JAK3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUP-Spleen;
RX MEDLINE=94192816; PubMed=8143863;
RA Takahashi T., Shirasawa T.;
RT "Molecular cloning of rat JAK3, a novel member of the JAK family of
RT protein tyrosine kinases.";
RL FEBS Lett. 342:124-128(1994).
CC -1- FUNCTION: TYROSINE KINASE OF THE NON-RECEPTOR TYPE, INVOLVED IN
```

```

CC THE INTERLEUKIN-2 AND INTERLEUKIN-4 SIGNALING PATHWAY.
CC PHOSPHORYLATES STAT6, IRS1, IRS2 AND PI3K.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: WHOLLY INTRACELLULAR, POSSIBLY MEMBRANE
CC ASSOCIATED (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TRANSCRIBED IN A VARIETY OF TISSUES INCLUDING
CC SPLEEN, LUNG, KIDNEY AND INTESTINE.
CC -1- DOMAIN: POSSESSES TWO PHOSPHOTRANSFERASE DOMAINS. THE SECOND ONE
CC PROBABLY CONTAINS THE CATALYTIC DOMAIN (BY SIMILARITY), WHILE THE
CC PRESENCE OF SLIGHT DIFFERENCES SUGGEST A DIFFERENT ROLE FOR
CC DOMAIN 1.
CC -1- PTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IL-2 AND IL-4 (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. JAK
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D28508; BAA05868.1; -
CC HSSP: P11362; IFGI.
CC InterPro: IPR000299; Band_4.1.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR000980; SH2.
CC InterPro: IPR001245; Tyr_kin.
CC Pfam: PF00069; pkinase; 2.
CC SMART: SM00295; B41; 1.
CC SMART: SM00252; SH2; 1.
CC SMART: SM00219; TyrKc; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 2.
CC PROSITE: PS50001; SH2; FALSE_NEG.
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW SH2 domain; Repeat.
KW SH2 DOMAIN 372 472 SH2 (ATYPICAL).
FT DOMAIN 517 777 PROTEIN KINASE 1.
FT DOMAIN 818 1091 PROTEIN KINASE 2.
FT NP_BIND 824 832 ATP (BY SIMILARITY).
FT BINDING 851 851 ATP (BY SIMILARITY).
FT ACT_SITE 945 945 BY SIMILARITY.
FT MOD_RES 976 976 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 1100 AA; 122560 MW; 1D59CA05F4DD7EE2 CRC64;

Query Match 44.08; Score 40; DB 1; Length 1100;
Best Local Similarity 57.18; Pred. No. 48;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 FRVSRYGSIYRRSR 17
| | | | | | | |
Db 878 FIVKRGVSYGPR 891

RESULT 11
JAK3_HUMAN
ID JAK3_HUMAN STANDARD; PRT; 1124 AA.
AC P52333; Q13259; Q13260; Q13611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TYROSINE-PROTEIN KINASE JAK3 (EC 2.7.1.112) (JANUS KINASE 3) (JAK-3)
DE (LEUCOCYTE JANUS KINASE) (L-JAK).
GN JAK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
```

Wed Feb 13 07:52:27 2002

us-09-485-571-27.rsp

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=94294384; PubMed=8022790;  
 RX Kawamura M., McVicar D.W., Johnston J.A., Blake T.B., Chen Y.-Q.,  
 RA Lal B.K., Lloyd A.R., Kelvin D.J., Staples J.E., Ortaldo J.R.,  
 RA O'Shea J.J.;  
 RA "Molecular cloning of L-JAK, a Janus family protein-tyrosine kinase  
 RT expressed in natural killer cells and activated leukocytes.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 91:6374-6378(1994).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=96027605; PubMed=7559633;  
 RX Lai K.S., Jin Y., Graham D.K., Witthuhn B.A., Ihle J.N., Liu E.T.;  
 RA "A kinase-deficient splice variant of the human JAK3 is expressed in  
 RT hematopoietic and epithelial cancer cells.";  
 RT J. Biol. Chem. 270:25028-25036(1995).  
 RL [3]  
 RN SEQUENCE OF 36-191 FROM N.A.  
 RP MEDLINE=96278845; PubMed=8662778;  
 RX Verbsky J.W., Bach E.A., Fang Y.F., Yang L., Randolph D.A.,  
 RA Fields L.E.;  
 RA "Expression of Janus kinase 3 in human endothelial and other non-  
 RT lymphoid and non-myeloid cells.";  
 RT J. Biol. Chem. 271:13976-13980(1996).  
 RL [4]  
 RN VARIANT SCID CYS-100.  
 RP MEDLINE=95388142; PubMed=7659163;  
 RX Macchi P., Villa A., Gillani S., Sacco M.G., Frattini A., Porta F.,  
 RA Ugazio A.G., Johnston J.A., Candotti F., O'Shea J.J., Vezzone P.,  
 RA Notarangelo L.D.;  
 RA "Mutations of Jak-3 gene in patients with autosomal severe combined  
 RT immune deficiency (SCID).";  
 RT Nature 377:65-68(1995).  
 RL [5]  
 RN VARIANTS SCID GLY-481; LEU-586--MET-592 DEL AND ARG-759.  
 RP MEDLINE=98022793; PubMed=9354668;  
 RX Candotti F., Oakes S., Johnston J.A., Gillani S., Schumacher R.F.,  
 RA Mella P., Fiorini M., Ugazio A.G., Badolato R., Notarangelo L.D.,  
 RA Bozzi F., Macchi P., Strina D., Vezzone P., Blaese R.M., O'Shea J.J.,  
 RA Villa A.;  
 RA "Structural and functional basis for JAK3-deficient severe combined  
 RT immunodeficiency.";  
 RT Blood 90:3996-4003(1997).  
 RL [6]  
 RN VARIANT SCID TRP-582.  
 RP MEDLINE=98423994; PubMed=9753072;  
 RX Bozzi F., Lefranc G., Villa A., Badolato R., Schumacher R.F.,  
 RA Khalil G., Lofset J., Bresciani S., O'Shea J.J., Vezzone P.,  
 RA Notarangelo L.D., Candotti F.;  
 RA "Molecular and biochemical characterization of JAK3 deficiency in a  
 RT patient with severe combined immunodeficiency over 20 years after  
 RT bone marrow transplantation: implications for treatment.";  
 RT Br. J. Haematol. 102:1363-1366(1998).  
 RL [7]  
 RN VARIANTS SCID ARG-151; ILE-722 AND SER-910.  
 RP MEDLINE=20435064; PubMed=10982185;  
 RX Schumacher R.F., Mella P., Badolato R., Fiorini M., Savoldi G.,  
 RA Giliani S., Villa A., Candotti F., Tampalini A., O'Shea J.J.,  
 RA Notarangelo L.D.;  
 RA "Complete genomic organization of the human JAK3 gene and mutation  
 RT analysis in severe combined immunodeficiency by single-strand  
 RT conformation polymorphism.";  
 RT Hum. Genet. 106:73-79(2000).  
 RL CC -!- FUNCTION: TYROSINE KINASE OF THE NON-RECEPTOR TYPE, INVOLVED IN  
 CC THE INTERLEUKIN-2 AND INTERLEUKIN-4 SIGNALING PATHWAY.  
 CC PHOSPHORYLATES STAT6, IRS1, IRS2 AND PI3K.  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: WHOLLY INTRACELLULAR, POSSIBLY MEMBRANE  
 CC ASSOCIATED (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: THREE SPLICED VARIANTS WERE ISOLATED FROM

FT FT VARIANT 586 592 /FTId-VAR\_010494.  
MISSING (IN SCID); LACK OF PHOSPHORYLATION  
IN RESPONSE TO CYTOKINE STIMULATION).  
FT FT VARIANT 722 722 /FTId-VAR\_010495.  
V -> I (IN SCID).  
FT FT VARIANT 759 759 /FTId-VAR\_010496.  
C -> R (IN SCID); CONSTITUTIVE  
PHOSPHORYLATION).  
FT FT VARIANT 910 910 /FTId-VAR\_010497.  
L -> S (IN SCID).  
FT FT VARIANT 34 34 /FTId-VAR\_010498.  
A -> G (IN REF. 2).  
FT FT CONFLICT 147 147 MISSING (IN REF. 3).  
FT FT CONFLICT 187 187 T -> A (IN REF. 3).  
FT FT CONFLICT 212 212 A -> R (IN REF. 2).  
FT FT CONFLICT 222 222 P -> R (IN REF. 2).  
FT FT CONFLICT 610 610 M -> I (IN REF. 2).  
FT FT CONFLICT 845 845 AH -> GD (IN REF. 2).  
FT FT CONFLICT 896 897 PE -> QS (IN REF. 2).  
SQ SEQUENCE 1124 AA; 125015 MW; 1D0FD22068E088E4 CRC64;

Query Match 44.0%; Score 40; DB 1; Length 1124;  
Best Local Similarity 57.1%; Pred. No. 49;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 FRVSYRGISYRRSR 17  
I I I I I I I  
Db 882 FIVKRGVSGYGPGR 895

RESULT 12  
JAK3\_MOUSE STANDARD; PRT; 1299 AA.  
ID JAK3\_MOUSE Q62137; Q61747; Q61746;  
AC Q62137; Q61747; Q61746;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TYROSINE-PROTEIN KINASE JAK3 (EC 2.7.1.112) (JANUS KINASE 3) (JAK-3).  
GN JAK3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Rodentia; Muridae; Sclurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=94309920; PubMed=7518579;  
RA Rane S.G., Reddy E.P.;  
RT "JAK3: a novel JAK kinase associated with terminal differentiation of  
hematopoietic cells.";  
RL Oncogene 9:2415-2423(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC STRAIN=BALB/C X 129 F2; TISSUE=Thymus;  
RX MEDLINE=96184772; PubMed=8605329;  
RA Gurniak C.B., Berg L.J.;  
RT "Murine JAK3 is preferentially expressed in hematopoietic tissues and  
lymphocyte precursor cells.";  
RL Blood 87:3151-3160(1996).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC STRAIN=BALB/C;  
RX MEDLINE=94294024; PubMed=8022486;  
RA Witthuhn B.A., Silvennoinen O., Miura O., Lai K.S., Cwik C., Liu E.T.,  
RA Ihle J.N.;  
RT "Involvement of the Jak-3 Janus kinase in signalling by interleukins  
2 and 4 in lymphoid and myeloid cells.";  
RL Nature 370:153-157(1994).  
CC -1- FUNCTION: TYROSINE KINASE OF THE NON-RECEPTOR TYPE, INVOLVED IN  
THE INTERLEUKIN-2 AND INTERLEUKIN-4 SIGNALING PATHWAY.  
CC PHOSPHORYLATES STAT6, IRS1, IRS2 AND PI3K.  
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
PROTEIN TYROSINE PHOSPHATE.

CC -1- SUBCELLULAR LOCATION: WHOLLY INTRACELLULAR, POSSIBLY MEMBRANE  
ASSOCIATED (BY SIMILARITY).  
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE  
PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN THE THYMUS WITH  
SOMEWAT LOWER LEVELS IN BONE MARROW, SPLEEN, FETAL LIVER AND  
ADULT CD4-CD8- THYMOCYTES. VERY LOW LEVELS IN ADULT KIDNEY, LUNG,  
TESTES, BRAIN AND LIVER.  
CC -1- DOMAIN: POSSESSES TWO PHOSPHOTRANSFERASE DOMAINS. THE SECOND ONE  
PROBABLY CONTAINS THE CATALYTIC DOMAIN (BY SIMILARITY), WHILE THE  
PRESENCE OF SLIGHT DIFFERENCES SUGGEST A DIFFERENT ROLE FOR  
DOMAIN 1.  
CC -1- FTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IL-2 AND IL-4.  
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. JAK  
SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L33768; AAA21415.1; -;  
CC EMBL; L40172; AAC42085.1; -;  
CC EMBL; L32955; AAA21565.1; -;  
CC HSSP; P11362; IFGI..  
CC MGD; MGI:99928; Jak3.  
CC InterPro: IPR000299; Band\_4.1.  
CC InterPro: IPR000719; Euk\_pkinase.  
CC InterPro: IPR000980; SH2.  
CC InterPro: IPR001245; Tyr\_kin.  
CC Pfam; PF00069; pkinase; 3.  
CC SMART; SM00295; B41; 1.  
CC SMART; SM00252; SH2; 1.  
CC SMART; SM00219; TyrKc; 1.  
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
CC PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 2.  
CC PROSITE; PS50001; SH2; FALSE\_NEG.  
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW SH2 domain; Repeat; Alternative splicing.  
FT DOMAIN 560 688 SH2 (ATYPICAL).  
FT DOMAIN 732 992 PROTEIN KINASE 1.  
FT DOMAIN 1032 1299 PROTEIN KINASE 2.  
FT NP\_BIND 1038 1046 ATP (BY SIMILARITY).  
FT BINDING 1064 1064 ATP (BY SIMILARITY).  
FT ACT\_SITE 1158 1158 BY SIMILARITY.  
FT MOD\_RES 1189 1189 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT VARSP\_LIC 222 238 RVWSPARPTATAGQVY -> LCGRLPGRPYALMAKYI  
(IN ISOFORM 3).  
FT VARSP\_LIC 223 238 VWSPARPTATAGQVY -> RVVACQADRYI (IN  
ISOFORM 2).  
FT VARSP\_LIC 284 473 MISSING (IN ISOFORM 2 AND ISOFORM 3).  
FT VARSP\_LIC 491 498 OPTCGSGR -> QAPRVGPAG (IN ISOFORM 2 AND  
ISOFORM 3).  
FT VARSP\_LIC 568 596 MISSING (IN ISOFORM 2 AND ISOFORM 3).  
FT VARSP\_LIC 656 673 ASASPTACGSCQLLEF -> GLSQPHRSRLRELLAACWNS  
(IN ISOFORM 2 AND ISOFORM 3).  
FT VARSP\_LIC 1000 1027 OTHPLASRVLESCAWPALCLPGRPHI -> SDPTPGIPS  
PRELVCVAGLAQLACQDPALF (IN ISOFORM 2 AND  
ISOFORM 3).  
FT VARSP\_LIC 1290 1299 SRTTGQPSAP -> EPHDRPAFATLSQDPLMLWRGPG  
(IN ISOFORM 2 AND ISOFORM 3).  
FT MUTAGEN L->R: LOSS OF ACTIVITY.  
FT CONFLICT 62 62 A -> G (IN REF. 2).  
FT CONFLICT 276 276 S -> P (IN REF. 3).  
FT CONFLICT 280 280 G -> N (IN REF. 3).  
FT CONFLICT 282 282 MISSING (IN REF. 3).  
FT CONFLICT 490 490 K -> N (IN REF. 3).  
FT CONFLICT 550 550 A -> P (IN REF. 2 AND 3).

Wed Feb 13 07:52:27 2002

```

FT CONFLICT 559 560 EL -> DV (IN REF. 3).
FT CONFLICT 607 607 A -> G (IN REF. 2 AND 3).
FT CONFLICT 683 683 N -> Y (IN REF. 3).
FT CONFLICT 706 706 T -> N (IN REF. 3).
FT CONFLICT 734 734 G -> EW (IN REF. 2 AND 3).
FT CONFLICT 750 750 R -> S (IN REF. 3).
FT CONFLICT 931 932 SG -> QR (IN REF. 3).
SQ SEQUENCE 1299 AA; 144314 MW; EFE2D60B6AF3D10C CRC64;

Query Match 44.0%; Score 40; DB 1; Length 1299;
Best Local Similarity 57.1%; Pred. No. 57;
Matches 8; Conservative 1; Mismatches 0; Indels 5; Gaps 0;

QY 4 FRVYRGISYRRS 17
Db 1091 FIVKRGVSGPCR 1104

RESULT 13
URED_ACTNA STANDARD; PRT; 271 AA.
AC Q92364;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UREASE ACCESSORY PROTEIN URED.
GN URED.
OS Actinomycetes naeslundii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VVU45;
RX MEDLINE=99115518; PubMed=9916052;
RA Morou-Bermudez E., Burne R.A.;
RT "Genetic and physiologic characterization of urease of Actinomycetes naeslundii."
RL Infect. Immun. 67:504-512(1999).
CC -!- FUNCTION: PROBABLY FACILITATES NICKEL INCORPORATION.
CC -!- SIMILARITY: BELONGS TO THE URED FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021815; CAA16996.1;
CC HSSP; P53608; 2CEV.
CC InterPro; IPR000287; Arginase.
CC Pfam; PF00491; arginase; 2.
CC PROSITE; PS00147; ARGINASE_1; FALSE_NEG.
CC PROSITE; PS00148; ARGINASE_2; 1.
CC PROSITE; PS01053; ARGINASE_3; 1.
CC Hypothetical protein; Hydrolase; Manganese; Signal; Multigene family.
KW SIGNAL
FT CHAIN 23 413 PUTATIVE AGMATINASE 2.
FT METAL 206 206 MANGANESE 1 (BY SIMILARITY).
FT METAL 229 229 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 231 231 MANGANESE 2 (BY SIMILARITY).
FT METAL 233 233 MANGANESE 1 (BY SIMILARITY).
FT METAL 331 331 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 333 333 MANGANESE 2 (BY SIMILARITY).
SQ SEQUENCE 413 AA; 45901 MW; 577EF7AF936E1346 CRC64;

Query Match 42.9%; Score 39; DB 1; Length 413;
Best Local Similarity 58.3%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGIS 12
Db 60 RWEFDYQYSGIS 71

RESULT 15
BCAL_RAT STANDARD; PRT; 968 AA.
ID BCAL_RAT
AC Q63767; Q63766;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DT CRK-ASSOCIATED SUBSTRATE (P130CAS) (BREAST CANCER ANTI-ESTROGEN RESISTANCE 1 PROTEIN).
DE RESISTANCE 1 PROTEIN.
GN BCAR1 OR CRKAS OR CAS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=94349922; PubMed=8070403;

Query Match 42.9%; Score 39; DB 1; Length 271;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 VSYRGISYRRS 16
Db 112 IAYRGASYRQT 122

RESULT 14
SPEC_SCHPO STANDARD; PRT; 413 AA.
ID SPEC_SCHPO
AC O42887;
DT 20-AUG-2001 (Rel. 40, Created)

```

RA Sakai R., Iwamatsu A., Hirano N., Ogawa S., Tanaka T., Mano H.,  
RA Yazaki Y., Hirai H.;  
RT "A novel signaling molecule, p130, forms stable complexes in vivo with  
RT v-Crk and v-Src in a tyrosine phosphorylation-dependent manner.";  
RL EMO J. 13:3748-3756(1994).  
RN [2]  
RP TYROSINE PHOSPHORYLATION BY FOCAL ADHESION KINASE.  
RX MEDLINE=96030588; PubMed=9360983;  
RA Tachibana K., Urano T., Fujita H., Ohashi Y., Kamiguchi K., Iwata S.,  
RA Hirai H., Morimoto C.;  
RT "Tyrosine phosphorylation of Crk-associated substrates by focal  
RT adhesion kinase. A putative mechanism for the integrin-mediated  
RT tyrosine phosphorylation of Crk-associated substrates.";  
RL J. Biol. Chem. 272:29083-29090(1997).  
CC -1- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE  
CC FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION.  
CC IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY).  
CC -1- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1.  
CC ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL.  
CC -1- SUBCELLULAR LOCATION: FOCAL ADHESIONS AND STRESS FIBERS.  
CC UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO  
CC THE MEMBRANE UPON TYROSINE PHOSPHORYLATION.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)  
CC AND A SHORT ISOFORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHER EXPRESSION IN LUNG,  
CC INTESTINE AND TESTIS.  
CC -1- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING  
CC MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN  
CC CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-  
CC BINDING SITES PUTATIVELY BIND CRK, NCK AND ABL SH2 DOMAINS. THE  
CC HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOPHYAL  
CC GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.  
CC -1- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM  
CC RESPONSE ELEMENT (SRE).  
CC -1- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE  
CC YDYLVL MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE  
CC PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE  
CC RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN  
CC MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX.  
CC -1- DISEASE: APPEARS TO HAVE A CENTRAL FUNCTION IN TRANSFORMATION OF  
CC SOME CELL TYPES.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN. NECESSARY FOR THE LOCALISATION  
CC OF THE PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-  
CC RICH REGION OF FOCAL ADHESION KINASE 1.  
CC -1- SIMILARITY: CONTAINS 1 SH3-BINDING DOMAIN. THE SH3 BINDING DOMAIN  
CC BINDS TO THE SRC SH3 DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE CAS FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; D29766; BAA06169.1; -;  
DR EMBL; D29766; BAA06170.1; -;  
DR HSSP; P29354; IGRI.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00002; SH3; 1.  
KW Phosphorylation; SH3 domain; SH3-binding; Cell adhesion;  
KW Alternative splicing  
FT DOMAIN 97 159  
FT DOMAIN 168 181  
FT DOMAIN 213 514  
FT DOMAIN 520 712  
FT DOMAIN 733 741  
FT DOMAIN 844 894  
FT VARSPLIC 5 98  
PRO-RICH.  
SUBSTRATE FOR KINASES.  
SER-RICH.  
SH3-BINDING (POTENTIAL).  
DIVERGENT HELIX-LOOP-HELIX MOTIF.  
MISSING (IN SHORT ISOFORM).

SQ SEQUENCE 968 AA; 104262 MW; E861641BFD68D377 CRC64;  
Query Match 42.9%; Score 39; DB 1; Length 968;  
Best Local Similarity 56.2%; Pred. No. 62;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
Qy 2 WSRVSVYRGISYRRSR 17  
Db 23 WGRVSVRRPQSYRAAR 38  
I III I III I  
I III I III I

Search completed: February 12, 2002, 12:39:53  
Job time: 806 sec

us-09-485-571-27.rsp

Wed Feb 13 07:52:27 2002

-----



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:41 ; Search time 232.64 Seconds  
(without alignments)  
10.689 Million cell updates/sec

Title: US-09-485-571-27

Perfect score: 91  
Sequence: 1 RWSFRVSYRGISYR 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	48	52.7	570	10 Q9LQE3	Q9LQE3 arabidopsis
2	47	51.6	615	10 Q9C719	Q9C719 arabidopsis
3	44.5	48.9	88	2 Q9KMK7	Q9KMK7 vibrio chol
4	44	48.4	620	10 Q9LHK2	Q9LHK2 arabidopsis
5	44	48.4	620	10 Q9C8N9	Q9C8N9 arabidopsis
6	43	47.3	660	5 Q17248	Q17248 boophilus m
7	43	47.3	1106	13 Q42291	Q42291 gallus gall
8	42	46.2	182	2 Q9RBT9	Q9RBT9 burkholderi
9	42	46.2	318	2 Q53813	Q53813 shigella bo
10	42	46.2	480	2 Q9RSL6	Q9RSL6 deinococcus
11	41	45.1	93	7 Q9TNR7	Q9TNR7 homo sapien
12	41	45.1	94	7 Q29785	Q29785 homo sapien
13	41	45.1	229	7 Q30079	Q30079 homo sapien
14	41	45.1	229	7 Q30080	Q30080 homo sapien
15	41	45.1	237	7 Q30097	Q30097 homo sapien
16	41	45.1	250	7 Q19709	Q19709 homo sapien
17	41	45.1	279	10 Q80901	Q80901 arabidopsis
18	41	45.1	307	1 Q9HSY4	Q9HSY4 halobacteri
19	41	45.1	323	5 Q9VA08	Q9VA08 drosophila

20	41	45.1	378	1	Q9HSZ6	Q9HSZ6 halobacteri
21	41	45.1	378	2	Q9AJU2	Q9AJU2 streptomyce
22	41	45.1	770	5	Q20908	Q20908 caenorhabdi
23	41	45.1	883	10	Q9M9X1	Q9M9X1 arabidopsis
24	41	45.1	949	5	Q9GPI1	Q9GPI1 heterodera
25	40	44.0	52	9	Q9MBU4	Q9MBU4 chlamydia p
26	40	44.0	89	7	Q98201	Q98201 homo sapien
27	40	44.0	90	7	Q29883	Q29883 homo sapien
28	40	44.0	106	7	Q30202	Q30202 homo sapien
29	40	44.0	105	7	Q9T076	Q9T076 homo sapien
30	40	44.0	108	7	Q90935	Q90935 homo sapien
31	40	44.0	146	2	Q9X507	Q9X507 streptomyce
32	40	44.0	229	7	Q29877	Q29877 homo sapien
33	40	44.0	229	7	Q30084	Q30084 homo sapien
34	40	44.0	250	7	Q19764	Q19764 homo sapien
35	40	44.0	261	7	Q30093	Q30093 homo sapien
36	40	44.0	261	7	Q29884	Q29884 homo sapien
37	40	44.0	306	11	Q9JUJ7	Q9JUJ7 rattus norv
38	40	44.0	330	1	Q9YBP0	Q9YBP0 aeropyrum p
39	40	44.0	347	11	Q9WUX9	Q9WUX9 rattus norv
40	40	44.0	366	7	P79587	P79587 rattus norv
41	40	44.0	378	2	Q9K722	Q9K722 bacillus ha
42	40	44.0	583	3	Q9C2K0	Q9C2K0 neurospora
43	40	44.0	713	2	Q9FC87	Q9FC87 streptomyce
44	40	44.0	781	5	Q9NF24	Q9NF24 caenorhabdi
45	40	44.0	821	10	Q9ZS31	Q9ZS31 solanum tub

#### ALIGNMENTS

RESULT 1  
Q9LQE3 PRELIMINARY; PRT: 570 AA.  
ID Q9LQE3  
AC Q9LQE3;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE F1504.42.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk S., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsy N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaver A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC007887; AAF79360.1;  
DR InterPro: IPR003340; B3.  
DR Pfam: PF02362; B3; 1.  
SQ SEQUENCE 570 AA; 65041 MW; E64A987BA713DF7E CRC64;

Query Match 52.7%; Score 48; DB 10; Length 570;  
Best Local Similarity 64.3%; Pred. No. 8.2;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RWSFRVSYRGISYR 14

STRAIN=EL TOR N16961 / SEROTYPE O1.  
MEDLINE=20406833; PubMed=10952301;  
Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
Fraser C.M.;  
"DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
*cholerae*.";  
Nature 406:477-483(2000).  
EMBL; AF004371; AAF96240.1; -;  
TIGR; VCA0332; -;  
Complete proteome.  
SEQUENCE 88 AA; 10092 MW; BI73D34A34B6A870 CRC64;  
KW

Query Match 48.9%; Score 44.5; DB 2; Length 88;  
Best Local Similarity 52.4%; Pred. No. 4.1;  
Matches 11; Conservative 2; Mismatches 3; Indels 5; Gaps 1;  
QY 2 WSGFRVSYRG-----ISYRRSR 17  
Db 58 WSGVITRGNTIRIISVRRSR 78

RESULT 4  
Q9LNK2 PRELIMINARY; PRT; 620 AA.  
ID Q9LNK2  
AC Q9LNK2  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE F12K21.26.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OC NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,  
Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,  
Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N., Nguyen M.,  
Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
Ecker J.R.;  
RA "Genomic sequence for Arabidopsis thaliana BAC F12K21 from chromosome  
I.";  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
Conn L., Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B.,  
Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N.,  
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
Thaveri A., Toriumi M., Vaysberg M., Yu G., Federspiel N.,  
Theologis A., Ecker J.;  
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AC023279; AAF79263.1; -;  
DR InterPro; IPR003340; B3.  
DR Pfam; PF02362; B3; 1.  
DR

STRAIN=CV. COLUMBIA;  
SEQUENCE 21016719; PubMed=11130712;  
RC Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,  
Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
Pai G., Peterson S.L., Rizzo P.K., Rizzio M., Rooney T., Rowley D.,  
Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
"Sequence and analysis of chromosome 1 of the plant Arabidopsis  
thaliana.";  
RT Nature 408:816-820(2000).  
RL EMBL; AC069160; AAG51458.1; -;  
DR InterPro; IPR003311; AUX.1AA.  
DR InterPro; IPR003340; B3.  
DR Pfam; PF02309; AUX.1AA; 1.  
DR Pfam; PF02362; B3; 1.  
KW Hypothetical protein.  
KW SEQUENCE 615 AA; 69755 MW; 817E7D03190622F7 CRC64;  
QY 1 RMSFRVSYRGISYRRS 16  
Db 175 QWRFRHSYRGTPQRHS 190

Query Match 51.6%; Score 47; DB 10; Length 615;  
Best Local Similarity 56.2%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
QY 175 QWRFRHSYRGTPQRHS 190  
Db

RESULT 3  
Q9KMK7 PRELIMINARY; PRT; 88 AA.  
ID Q9KMK7  
AC Q9KMK7  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE HYPOTHETICAL PROTEIN VCA0332.  
GN VCA0332.  
OS *Vibrio cholerae*.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.  
OC NCBI\_TaxID=666;  
DR InterPro; IPR003340; B3.  
DR Pfam; PF02362; B3; 1.  
DR

SO SEQUENCE 620 AA; 70895 MW; C2FE2FAFDECC006DE CRC64;

Query Match 48.4%; Score 44; DB 10; Length 620;  
Best Local Similarity 50.0%; Pred. No. 42;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 RWSFRVSYRGISYRRS 16  
: ||| :||| :|||  
Db 150 QWRFRHNYRGTPQRHS 165

RESULT 5  
ID Q9C8N9 PRELIMINARY; PRT; 620 AA.

AC Q9C8N9;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE AUXIN RESPONSE FACTOR, PUTATIVE.  
GN F7P12.12.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]

SEQUENCE FROM N.A.

RC STRAIN-CV, COLUMBIA;  
RX MEDLINE=21016719; PubMed=11130712;  
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
Buehler E., Chao A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kvan A., Lam B.,  
Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
Millschler J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
Uttarback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
thaliana.";  
RL Nature 408:816-820(2000).  
DR EMBL; AC023913; AAG51897.1; -.  
DR InterPro; IPR003311; AUX\_1AA.  
DR InterPro; IPR003340; B3.  
DR Pfam; PF02309; AUX\_1AA; 1.  
DR Pfam; PF02362; B3; 1.  
SQ SEQUENCE 620 AA; 70555 MW; 88484FF185EBED3F CRC64;

Query Match 48.4%; Score 44; DB 10; Length 620;  
Best Local Similarity 50.0%; Pred. No. 42;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 RWSFRVSYRGISYRRS 16  
: ||| :||| :|||  
Db 187 QWRFRHNYRGTPQRHS 202

RESULT 6  
Q17248 PRELIMINARY; PRT; 660 AA.

AC Q17248;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE ANGIOTENSIN-CONVERTING ENZYME-LIKE PROTEIN PRECURSOR.  
GN BM91.  
OS Boophilus microplus (Cattle tick).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.  
OX NCBI\_TaxID=6941;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-WHOLE TICKS;  
RA Whitfield P.L., Johnson M.C., Smith D.R.J., Snelson V.J., Clifton G.F.,  
Brown G.S., Cairns D., Foy A.B., Irving D.O.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U62809; AAB04998.1; -.  
DR InterPro; IPR001548; Peptidase\_M2.  
DR Pfam; PF01401; Peptidase\_M2; 1.  
DR PRINTS; PR00791; PEPDIPTASEA.  
DR PRODOM; PD004184; Peptidase\_M2; 1.  
KW Signal.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 660 POTENTIAL.  
SQ SEQUENCE 660 AA; 75257 MW; 6F164CF70C938E63 CRC64;

Query Match 47.3%; Score 43; DB 5; Length 660;  
Best Local Similarity 45.5%; Pred. No. 65;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 WSRFVSRYGIS 12  
: ||| :||| :|||  
Db 488 WEYRIKQGV 498

RESULT 7  
O42291 PRELIMINARY; PRT; 1106 AA.

ID O42291  
AC O42291;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE JANUS TYROSINE KINASE.  
GN JAK.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SPLEEN;  
RA Sofer L., Kampa D., Burnside J.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF034576; AAC34195.1; -.  
DR HSP; P12931; IFMK.  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR InterPro; IPR000980; SH2.  
DR InterPro; IPR001245; Tyr\_Kin.  
DR Pfam; PF00069; pkinase; 2.  
DR Pfam; PF00017; SH2; 1.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00252; SH2; 1.  
DR SMART; SM00219; TyrcK; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 2.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50001; SH2; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1106 AA; 124750 MW; 004F75F851A282B1 CRC64;

Query Match 47.3%; Score 43; DB 13; Length 1106;  
Best Local Similarity 57.1%; Pred. No. 1.1e-02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Wed Feb 13 07:52:28 2002

RESULT 10  
QY 4 FRVSYRGISYRRSR 17  
Db 881 FIVKYRGVCSYGR 894

PRELIMINARY; PRT; 480 AA.

Q9RSL6  
AC Q9RSL6  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE MOLYBDATE METABOLISM REGULATOR-RELATED PROTEIN.  
GN DR2108.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
Fraser C.M.;  
RA "Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1".  
RL Science 286:1571-1577(1999).  
DR EMBL; AF002046; AAF11657.1;  
DR TIGR; DR2108;  
KW Complete proteome.  
SQ SEQUENCE 480 AA; 52430 MW; B67C9175AAC14281 CRC64;

Query Match 46.2%; Score 42; DB 2; Length 480;  
Best Local Similarity 50.0%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 WSRVSYRGISYRRSR 17  
Db 234 WMRVDARGVYGHRSR 249

RESULT 11  
QY 2 WSRVSYRGISYRRSR 17  
Db 234 WMRVDARGVYGHRSR 249

PRELIMINARY; PRT; 93 AA.

Q9TNR7  
AC Q9TNR7  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE HLA-DQB1\*0501 BETA CHAIN-MHC CLASS II ANTIGEN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93143796; PubMed=7678744;  
RA Olsen M.L., Annett F.C., Reveille J.D.;  
RT "Contrasting molecular patterns of MHC class II alleles associated  
with the anti-Sm and anti-RNP precipitin autoantibodies in systemic  
lupus erythematosus".  
RL Arthritis Rheum. 36:94-104(1993).  
CC -1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR  
HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,  
BETA-1 DOMAIN INTERPRO FAMILY.  
CC InterPro: IPR000353; MHC\_II\_beta.  
DR Pfam: PF00969; MHC\_II\_beta: 1.  
DR ProDom: PD000328; MHC\_II\_beta: 1.  
KW Glycoprotein; MHC; MHC II; Transmembrane.  
SQ SEQUENCE 93 AA; 10990 MW; A81833111D67962E CRC64;

QY 4 FRVSYRGISYRRSR 17  
Db 881 FIVKYRGVCSYGR 894

PRELIMINARY; PRT; 182 AA.

Q9RBT9  
AC Q9RBT9  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE P-HYDROXYLAMINO BENZOATE LYASE.  
GN PNB8.  
OS Burkholderia pickettii (Pseudomonas pickettii).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TH105;  
RA Yabannavar A., Zylstra G.J.;  
RT "Analysis of the genes for p-nitrobenzoate degradation from Ralstonia  
pickettii YH105".  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF187879; AAF01444.1;  
KW Lyase  
SQ SEQUENCE 182 AA; 20053 MW; 45FA1636CAC43E06 CRC64;

Query Match 46.2%; Score 42; DB 2; Length 182;  
Best Local Similarity 43.8%; Pred. No. 24;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 WSRVSYRGISYRRSR 17  
Db 63 WAAATIDIEGPNYRRSR 78

RESULT 9  
QY 2 WSRVSYRGISYRRSR 17  
Db 63 WAAATIDIEGPNYRRSR 78

PRELIMINARY; PRT; 318 AA.

Q53813  
AC Q53813  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)  
DE BY' GENE PRODUCT (FRAGMENT).  
OS Shigella boydii.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=621;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91286192; PubMed=2061288;  
RA Tominaga A., Ikemizu S., Enomoto M.;  
RT "Site-specific recombinase genes in three Shigella subgroups and  
nucleotide sequences of a pinB gene and an invertible B segment from  
Shigella boydii".  
RL J. Bacteriol. 173:4079-4087(1991).  
DR EMBL; D00660; BAA00555.1;  
FT NON\_TER 318 318  
SQ SEQUENCE 318 AA; 34335 MW; 93CC950314D4E275 CRC64;

Query Match 46.2%; Score 42; DB 2; Length 318;  
Best Local Similarity 62.5%; Pred. No. 43;  
Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 4 FRVSYRGISYRRSR 17  
Db 205 FRVNRNGGIFYRRSR 220

```

Query Match      45.1%; Score 41; DB 7; Length 93;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRR 15
DB 79 RHNYEVAYRILQRR 93

RESULT 12
ID Q29785 PRELIMINARY; PRT; 94 AA.
AC Q29785;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE MHC CLASS II HLA-DQ-BETA-1 FIRST DOMAIN MRNA (DRW6,DQW1.9), PARTIAL
DE CDS (FRAGMENT).
DE HLA-DQB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88145646; PubMed=2894075;
RA Sinha A.A., Brautbar C., Szafer F., Friedmann A., Tzfoni E.,
RA Todd J.A., Steinman L., McDewitt H.O.;
RT "A newly characterized HLA DQ beta allele associated with pemphigus
RT vulgaris."
RL Science 239:1026-1029(1988).
CC -1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR
CC HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,
CC BETA-1 DOMAIN INTERPRO FAMILY.
DR EMBL; M19239; AA52319.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
DR Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA; 11169 MW; 65EBB6D0400BED61 CRC64;

Query Match      45.1%; Score 41; DB 7; Length 94;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRR 15
DB 80 RHNYEVAYRILQRR 94

RESULT 13
ID Q30079 PRELIMINARY; PRT; 229 AA.
AC Q30079;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE MHC CLASS II HLA-DQ-BETA-1 (FRAGMENT).
DE HLA-DQB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97083137; PubMed=8929711;
RA Yasunaga S., Kimura A., Hamaguchi K., Ronningen K.S., Sasazuki T.;
RT "Different contribution of HLA-DR and -DQ genes in susceptibility and
RT resistance to insulin-dependent diabetes mellitus (IDDM).";

```

```

Tissue Antigens 47:37-48(1996).
-1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR
CC HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
CC -1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,
CC BETA-1 DOMAIN INTERPRO FAMILY.
DR EMBL; L34102; AAC41970.1; -.
DR HSSP; P06343; IIAK.
DR InterPro; IPR003597; Ig_cl.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IGcl; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 229 AA; 26318 MW; 64A5452E72F4BD45 CRC64;

Query Match      45.1%; Score 41; DB 7; Length 229;
Best Local Similarity 53.3%; Pred. No. 44;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 RWSFRVSYRGISYRR 15
DB 80 RHNYEVAYRILQRR 94

RESULT 14
ID Q30080 PRELIMINARY; PRT; 229 AA.
AC Q30080;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE MHC CLASS II HLA-DQ-BETA-1 (FRAGMENT).
DE HLA-DQB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97083137; PubMed=8929711;
RA Yasunaga S., Kimura A., Hamaguchi K., Ronningen K.S., Sasazuki T.;
RT "Different contribution of HLA-DR and -DQ genes in susceptibility and
RT resistance to insulin-dependent diabetes mellitus (IDDM).";
RL Tissue Antigens 47:37-48(1996).
CC -1- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR
CC HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
CC -1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,
CC BETA-1 DOMAIN INTERPRO FAMILY.
DR EMBL; L34103; AAC41971.1; -.
DR HSSP; P06343; IIAK.
DR InterPro; IPR003597; Ig_cl.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IGcl; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 229 AA; 26337 MW; B494C32F3D782F08 CRC64;

Query Match      45.1%; Score 41; DB 7; Length 229;

```

Wed Feb 13 07:52:28 2002

Best Local Similarity	53.3%	Pred. No. 44;	0;
Matches	8;	Conservative 3;	Mismatches 4;
		Indels	0;
		Gaps	0;

```

1 RWSFRVSYRGISYRR 15
  | :: |::| | |
80 RHNYEVAYRGILQRR 94

```

RESULT 15

30097/ PRELIMINARY; PRT; 237 AA.

Q30097;  
C 01-NOV-1996 (TReMBLrel. 01, Created)  
T 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
T 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
T 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
E MHC CLASS II HLA-DQ-BETA-1 (FRAGMENT).  
N HLA-DOB1.  
S Homo sapiens (Human).  
NS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
OX [1]  
XX SEQUENCE FROM N.A.  
Y Yasunaga S.;  
RA Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
C ~1- MISCELLANEOUS; FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR  
C HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).  
C ~1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
C DOMAIN.  
C ~1- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,  
C BETA-1 DOMAIN INTERPRO FAMILY.  
C EMBL; L40180; AAA92331.1; -.  
C HSSP; P06343; IIAK.  
C InterPro; IPR003597; Ig\_cl.  
C InterPro; IPR003006; Ig\_MHC.  
C InterPro; IPR000353; MHC\_II\_beta.  
C Pfam; PF00047; ig; 1.  
C Pfam; PF00969; MHC\_II\_beta; 1.  
C Prodom; PD000328; MHC\_II\_beta; 1.  
C SMART; SM00407; IGcl; 1.  
C PROSITE; PS00290; IG\_MHC; 1.  
C Alternative splicing; Glycoprotein; MHC; MHC II; Transmembrane.  
C NON\_TER 1  
C SEQUENCE 237 AA; 27038 MW; 460BG3ABCA7BC8EA CRC64;  
C SO

Query Match 45.1%; Score 41; DB 7; Length 237;  
Best Local Similarity 53.3%; Pred. No. 46;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

1 RWSFRVSYRGISYRR 15  
| :: | : | | | | | |  
80 RHNYEVAYRGILQRR 94

Search completed: February 12, 2002, 12:38:42  
Job time: 755 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:33 ; Search time 242.57 Seconds  
(without alignments)  
5.191 Million cell updates/sec

Title: US-09-485-571-30  
Perfect score: 91  
Sequence: 1 RSRYSIGRYSVRSWK 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	91	100.0	17	AAW99415	Tachyplesin deriva
2	44	48.4	19	AAW23770	Tachytegrin and/or
3	44	48.4	19	AAW23769	Tachytegrin and/or
4	42	46.2	66	AAW48001	Arabidopsis thalia
5	42	46.2	87	AAW48000	Arabidopsis thalia
6	41.5	45.6	3025	AAW86197	HIV-1 subtype C pr
7	41	45.1	717	AAW41819	Human OREF ORF1583
8	41	45.1	717	AAW39298	Human polypeptide
9	41	45.1	734	AAW39299	Human polypeptide
10	41	45.1	734	AAW94096	Human protein sequ
11	41	45.1	738	AAW41084	Human polypeptide

12	41	45.1	738	22	AAW41085	Human polypeptide
13	41	45.1	1048	21	AAW42154	Arabidopsis thalia
14	41	45.1	1129	21	AAW42153	Arabidopsis thalia
15	41	45.1	1193	21	AAW42152	Arabidopsis thalia
16	40	44.0	18	18	AAW233779	Tachytegrin and/or
17	40	44.0	18	18	AAW233766	Tachytegrin and/or
18	40	44.0	398	21	AAW233676	Arabidopsis thalia
19	40	44.0	412	21	AAW233675	Arabidopsis thalia
20	40	44.0	427	21	AAW97536	B. lactofermentum
21	40	44.0	427	21	AAW97537	B. lactofermentum
22	40	44.0	855	13	AAW28033	Env polypeptide of
23	40	44.0	855	15	AAW51251	FIV JAPANW2 envel
24	39	42.9	85	22	AAW63820	Human prostate can
25	38	41.8	14	22	AAW98794	Human cell death p
26	38	41.8	19	18	AAW23768	Tachytegrin and/or
27	38	41.8	19	18	AAW23780	Tachytegrin and/or
28	38	41.8	254	20	AAW25950	Human ICAM-2 prote
29	38	41.8	254	20	AAW25951	Human ICAM-2 prote
30	38	41.8	254	20	AAW25952	Human ICAM-2 prote
31	38	41.8	254	20	AAW25953	Human ICAM-2 prote
32	38	41.8	254	20	AAW25954	Chimp ICAM-2 prote
33	38	41.8	254	20	AAW25955	Chimp ICAM-2 prote
34	38	41.8	254	20	AAW25956	Gorilla ICAM-2 pro
35	38	41.8	275	22	AAW38913	Human polypeptide
36	38	41.8	283	22	AAW85041	Shrimp white spot
37	38	41.8	305	22	AAW40699	Human polypeptide
38	38	41.8	544	19	AAW46314	AAV4 VP3 coat prot
39	38	41.8	568	21	AAW76036	Rat skin cell prot
40	38	41.8	568	22	AAW55975	Skin cell protein,
41	38	41.8	598	19	AAW46313	AAV4 VP2 coat prot
42	38	41.8	734	19	AAW46308	AAV4 VP1 capsid pr
43	38	41.8	900	20	AAW96264	Human axin. Homo
44	38	41.8	1487	19	AAW76719	Thermococcus 9N2 b
45	38	41.8	1487	19	AAW76720	Thermococcus 9N2 m

ALIGNMENTS

RESULT 1	AAW99415	standard; peptide; 17 AA.
ID	AAW99415	standard; peptide; 17 AA.
XX	AC	AAW99415;
XX	DT	08-JUN-1999 (first entry)
XX	DE	Tachyplesin derivative peptide SW2310.
XX	DE	Linear: tachyplesin; peptide antibiotic; beta-sheet; secondary structure;
KW	KW	disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;
KW	KW	anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;
KW	KW	nucleus; blood-brain barrier.
XX	OS	Synthetic.
XX	OS	WO9907728-A2.
PN	XX	18-FEB-1999.
PD	XX	06-AUG-1998; 98WO-FR01757.
PF	XX	12-AUG-1997; 97FR-0010297.
PR	XX	(SYNT-) SYNT:EM SA.
PA	XX	Calas B, Chavanieu A, Grassy G, Kaczorek M;
PI	XX	WPI; 1999-190034/16.
DR	XX	Derivatives of antibiotic peptides lacking disulfide bridges - used
PT	XX	as carriers to deliver active agents into cells
XX	XX	

PS Claim 8; Page 28; 37pp; French.  
XX This peptide represents a linear derivative of the tachyplesin family of  
CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide  
CC antibiotic family which contain a beta-sheet secondary structure linked  
CC by disulphide bridges. The new derivatives are linear and lack the  
CC disulphide bridge. The novel derivatives are used to deliver active  
CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
CC including crossing the blood-brain barrier.  
XX  
SQ Sequence 17 AA;

Query Match 100.0%; Score 91; DB 20; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRYSIGRYSVRFSWK 17  
Db 1 rrrysigrvsvrfswk 17

RESULT 2  
AAW23770  
ID AAW23770 standard; peptide; 19 AA.  
XX  
AC AAW23770;  
XX  
DT 16-OCT-1997 (first entry)  
XX  
DE Tachytegrin and/or parevin compound peptide.  
XX  
KW Protegrin; tachyplesin; antimicrobial; antiviral; oral mucositis;  
KW sexually transmitted disease; endotoxin; HIV; Staphylococcus aureus;  
KW Pseudomonas; H. pylori; Chlamydia trachomatis; Treponema pallidum;  
KW Neisseria gonorrhoeae; Trichomonas vaginalis; Herpes simplex type 2;  
KW animal; plant.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Disulfide-bond 4..17  
FT /label= Optional  
FT Modified-site 8  
FT /label= MeCly  
FT Modified-site 13  
FT /label= MeCly  
FT Modified-site 19  
FT /note= "Optionally amidated"  
XX  
XX W09702287-Al.  
XX  
XX 23-JAN-1997.  
XX  
XX 03-JUL-1996; 96WO-US11323.  
XX  
XX 06-JUL-1995; 95US-0000898.  
XX  
XX (INTR-) INTRABIOTICS PHARM INC.  
XX  
XX Chang CC, Gu CL, Harwig SSL, Lehrer RI;  
XX WPI; 1997-108914/10.  
XX  
XX Cpd. based on protegrin(s) and tachyplesin(s) - designated  
XX tachytegrin(s) and parevin(s), useful as antimicrobial and antiviral  
XX agents, e.g. for treatment of sexually transmitted diseases  
XX  
XX Claim 21; Page 44; 59pp; English.  
PS

XX The present sequence represents a peptide sequence found in compounds  
CC based on protegrins and tachyplesins, designated tachytegrins and  
CC parevins. The present sequence is from a tachytegrin as the peptide has  
CC Cys residues at positions 4 and 17, where parevins have Cys residues at  
CC positions 5 and 16. Both compounds can be used as preservatives (e.g.  
CC in foodstuffs, cosmetics or medications) or in treatment of viral or  
CC microbial infection in animals or plants. They are useful, e.g. in  
CC treatment of oral mucositis, infection by Staphylococcus aureus,  
CC pseudomonas or H. pylori, or sexually transmitted diseases, including  
CC those caused by Chlamydia trachomatis, Treponema pallidum, Neisseria  
CC gonorrhoeae, Trichomonas vaginalis, Herpes simplex type 2 and HIV. DNA  
CC encoding the compounds can also be expressed in situ, in animals or  
CC plants, to combat infections. The compounds can also be used as  
CC standards in antimicrobial assays and in binding endotoxins.  
XX  
SQ Sequence 19 AA;

Query Match 48.4%; Score 44; DB 18; Length 19;  
Best Local Similarity 63.6%; Pred. No. 2;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 RYSIGRYSVRF 14  
Db 6 ryxvgrfxvrf 16

RESULT 3  
AAW23769  
ID AAW23769 standard; peptide; 19 AA.  
XX  
AC AAW23769;  
XX  
DT 16-OCT-1997 (first entry)  
XX  
DE Tachytegrin and/or parevin compound peptide.  
XX  
KW Protegrin; tachyplesin; antimicrobial; antiviral; oral mucositis;  
KW sexually transmitted disease; endotoxin; HIV; Staphylococcus aureus;  
KW Pseudomonas; H. pylori; Chlamydia trachomatis; Treponema pallidum;  
KW Neisseria gonorrhoeae; Trichomonas vaginalis; Herpes simplex type 2;  
KW animal; plant.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Disulfide-bond 4..17  
FT /label= Optional  
FT Modified-site 19  
FT /note= "Optionally amidated"  
XX  
XX W09702287-Al.  
XX  
XX 23-JAN-1997.  
XX  
XX 03-JUL-1996; 96WO-US11323.  
XX  
XX 06-JUL-1995; 95US-0000898.  
XX  
XX (INTR-) INTRABIOTICS PHARM INC.  
XX  
XX Chang CC, Gu CL, Harwig SSL, Lehrer RI;  
XX WPI; 1997-108914/10.  
XX  
XX Cpd. based on protegrin(s) and tachyplesin(s) - designated  
XX tachytegrin(s) and parevin(s), useful as antimicrobial and antiviral  
XX agents, e.g. for treatment of sexually transmitted diseases  
XX  
XX Claim 21; Page 44; 59pp; English.  
XX  
XX The present sequence represents a peptide sequence found in compounds  
CC



CC based on protegrins and tachyplesins, designated tachytegrins and  
 CC parevins. The present sequence is from a tachytegrin as the peptide has  
 CC Cys residues at positions 4 and 17, where parevins have Cys residues at  
 CC positions 5 and 16. Both compounds can be used as preservatives (e.g.  
 CC in foodstuffs, cosmetics or medicaments) or in treatment of viral or  
 CC microbial infection in animals or plants. They are useful, e.g. in  
 CC treatment of oral mucositis, infection by *Staphylococcus aureus*,  
 CC *Pseudomonas* or *H. pylori*, or sexually transmitted diseases, including  
 CC those caused by *Chlamydia trachomatis*, *Treponema pallidum*, *Neisseria*  
 CC gonorrhoeae, *Trichomonas vaginalis*, Herpes simplex type 2 and HIV. DNA  
 CC encoding the compounds can also be expressed *in situ*, in animals or  
 CC plants, to combat infections. The compounds can also be used as  
 CC standards in antimicrobial assays and in binding endotoxins.

XX Sequence 19 AA;

Query Match 48.4%; Score 44; DB 18; Length 19;  
 Best Local Similarity 63.6%; Pred. No. 2;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 RYSIGRYSVRF 14  
 ||:||||  
 Db 6 rykvgrfkvr 16

#### RESULT 4

AAG48001  
 ID AAG48001 standard; Protein: 56 AA.

XX AC AAG48001;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60568.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126284.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0132863.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.

PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155139.
PR	24-SEP-1999;	99US-0155486.
PR	28-SEP-1999;	99US-0155659.
PR	29-SEP-1999;	99US-0156458.
PR	04-OCT-1999;	99US-0156596.
PR	05-OCT-1999;	99US-0157117.
PR	06-OCT-1999;	99US-0157753.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161932.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 46.2%; Score 42; DB 21; Length 66;		
Best Local Similarity 46.2%; Pred. No. 14;		
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;		
QY	5	YSIGRYSVRFWSK 17
Db	37	fslgryfslpwk 49
:1:111:11		
RESULT 5		
AAG48000		
ID	AAG48000 standard; Protein; 87 AA.	
XX	AC AAG48000;	
XX	18-OCT-2000 (first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 60567.	
XX	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX	Arabidopsis thaliana.	
XX	EPI033405-A2.	
PD	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
PF	25-FEB-1999; 99US-0121825.	
XX	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	
PR	23-APR-1999; 99US-0130510.	
PR	23-APR-1999; 99US-0130891.	
PR	28-APR-1999; 99US-0131449.	
PR	30-APR-1999; 99US-0132048.	
PR	04-MAY-1999; 99US-0132407.	
PR	05-MAY-1999; 99US-0132484.	
PR	06-MAY-1999; 99US-0132485.	
PR	07-MAY-1999; 99US-0132486.	
PR	11-MAY-1999; 99US-0132487.	
PR	14-MAY-1999; 99US-0132863.	
PR	14-MAY-1999; 99US-0134256.	
PR	14-MAY-1999; 99US-0134218.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134221.	
PR	14-MAY-1999; 99US-0134370.	
PR	18-MAY-1999; 99US-0134768.	
PR	19-MAY-1999; 99US-0134941.	
PR	20-MAY-1999; 99US-0135124.	
PR	21-MAY-1999; 99US-0135353.	
PR	24-MAY-1999; 99US-0135629.	

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140333.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0155179.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-01611993.  
PR 29-OCT-1999; 99US-0162142.  
Query Match 46.2%; Score 42; DB 21; Length 87;  
Best Local Similarity 46.2%; Pred. No. 19;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 YSIGRYSVRFSSWK 17  
:|||||: ||  
DB 58 fslgryfslpwk 70

RESULT 6  
ID AAB86197 standard; Protein; 3025 AA.  
XX AAB86197;  
XX 21-AUG-2001 (first entry)  
XX HIV-1 subtype C protein fragment #3.  
XX Infection; diagnosis; human; humoral immune response; antiviral;  
KW cellular immune response; vaccine; treatment; gene therapy.  
XX Human immunodeficiency virus type 1.  
XX Key Location/Qualifiers  
FH Misc-difference 1..3025  
FT /label= Xaa  
FT /note= "Xaa represents a stop codon"  
XX DE10056747-A1.  
XX 31-MAY-2001.  
XX 16-NOV-2000; 2000DE-1056747.  
XX 16-NOV-1999; 99DE-1055089.  
XX (SHAO/) SHAO Y.  
PA (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.  
XX Wagner R, Wolf H, Shao Y, Graf M;  
XX WPI; 2001-336417/36.  
XX New nucleic acid sequences from a human immune deficiency virus  
PT intersubtype, useful for treatment, prevention and diagnosis of  
PT infection  
XX Disclosure; Fig 8A-O; 48pp; German.  
XX This invention describes a novel polynucleotide isolated from human  
CC immunodeficiency virus type 1 subtype C/B' which can be used for the  
CC induction of specific humoral and cellular immune responses. (I) and  
CC polypeptides (II) encoded by them, are useful in pharmaceuticals,  
CC vaccines and diagnostic agents, particularly for treatment or prevention  
CC of human immune deficiency virus-1 (HIV-1) infections, also for rational  
CC design of test or therapeutic reagents, or gene therapy vectors.  
CC Polypeptides, especially antibodies, specifically directed against (II)  
CC are similarly useful as pharmaceutical and diagnostic agents. (I) are  
CC specific for intersubtype C/B' of HIV-1 so are useful in regions  
CC (particularly China and South-East Asia) where this subtype is prevalent.  
CC The products of the invention have antiviral activity. This sequence  
CC represents a protein encoded by the HIV-1 subtype C genome described in  
CC the method of the invention.  
XX Sequence 3025 AA;

Query Match 45.6%; Score 41.5; DB 22; Length 3025;  
Best Local Similarity 50.0%; Pred. No. 8.1e+02;

Matches 9; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 RSRYSIGR-YSVRFSSWK 17  
:|||||: ||  
DB 607 dsyrysisrtyctqgnwk 624

RESULT 7  
AAB41819  
ID AAB41819 standard; Protein; 717 AA.  
XX AAB41819;  
XX 08-FEB-2001 (first entry)  
XX Human ORFX ORF1583 polypeptide sequence SEQ ID NO:3166.  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; hypothyroidism; SCID; AIDS;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus; infection;  
KW cholesterol ester storage; nocturnal haemoglobinuria; burn; wound;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX Homo sapiens.  
XX WO200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US08621.  
XX 31-MAR-1999; 99US-0127607.  
XX 02-APR-1999; 99US-0127636.  
XX 05-APR-1999; 99US-0127728.  
XX 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
XX Shinkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX N-PSDB; AAC76028.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 11; Page 2383-2384; 5507pp; English.  
XX AAC7446 to AAC77606 encode the proteins given in AABA0237 to AABA43397,  
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX  
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;  
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
XX anticonvulsant; thrombolytic; coagulant; immunosuppressant;  
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressive;  
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
XX antidiabetic; hypotensive; dermatological; immunosuppressive;  
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
XX antithyroid; and antianemic. The sequences can be used for determining  
XX the presence of or predisposition to, or preventing or treating  
XX pathological conditions associated with an ORFX-associated disorder. The  
XX nucleic acids can be used to express ORFX proteins in gene therapy  
XX vectors. The proteins and nucleic acids may be used to treat cancers,  
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,  
XX graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease: to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 SQ Sequence 717 AA;

Query Match 45.1%; Score 41; DB 21; Length 717;  
 Best Local Similarity 37.5%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRFSWK 17  
 :||:|:|:|:|:|:  
 Db 420 arrfsagweargrwr 435

RESULT 8  
 AAM39298  
 ID AAM39298 standard; Protein; 717 AA.

XX AC AAM39298;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 2443.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

XX DR N-PSDB; AAI58454.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Example 4; SEQ ID NO 2443; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX SQ Sequence 717 AA;

Query Match 45.1%; Score 41; DB 22; Length 717;  
 Best Local Similarity 37.5%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRFSWK 17  
 :||:|:|:|:|:|:  
 Db 420 arrfsagweargrwr 435

RESULT 9  
 AAM39299  
 ID AAM39299 standard; Protein; 734 AA.

XX AC AAM39299;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 2444.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

XX DR N-PSDB; AAI58455.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Example 4; SEQ ID NO 2444; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

Wed Feb 13 07:52:29 2002

us-09-485-571-30.rag

CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 734 AA;

Query Match 45.1%; Score 41; DB 22; Length 734;

Best Local Similarity 37.5%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRFSWK 17

Db 437 arrfsagqwearrgwr 452  
 :||:| |:: | |:

RESULT 10

AAB94096

ID AAB94096 standard; Protein; 734 AA.

XX AC AAB94096;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:14312.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 XX full-length cDNAs defined in the specification, and for the detection  
 XX and/or diagnosis of the abnormality of the proteins encoded by the  
 XX full-length cDNAs -

XX Claim 8; SEQ ID 14312; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602  
 XX full-length cDNAs defined in the specification. Where a primer set  
 XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 XX to the complementary strand of a polynucleotide which comprises one of  
 XX the 5602 nucleotide sequences defined in the specification, where the  
 XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences, and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 734 AA;

Query Match 45.1%; Score 41; DB 22; Length 734;

Best Local Similarity 37.5%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRFSWK 17

Db 437 arrfsagqwearrgwr 452  
 :||:| |:: | |:

RESULT 11

AAM41084

ID AAM41084 standard; Protein; 738 AA.

XX AC AAM41084;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 6015.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 XX peripheral nervous system; neuropathy; central nervous system; CNS;  
 XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 XX leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60240.

XX Novel nucleic acids and polypeptides, useful for treating disorders



us-09-485-571-30.rag

Wed Feb 13 07:52:29 2002

PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	22-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	23-JUL-1999;	99US-0145276.
PR	26-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157153.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.



PR	18-OCT-1999;	9905-0159584.	PR	07-MAY-1999;	9905-0132863.
PR	21-OCT-1999;	9905-0160741.	PR	11-MAY-1999;	9905-0134256.
PR	21-OCT-1999;	9905-0160767.	PR	14-MAY-1999;	9905-0134218.
PR	21-OCT-1999;	9905-0160768.	PR	14-MAY-1999;	9905-0134219.
PR	21-OCT-1999;	9905-0160770.	PR	14-MAY-1999;	9905-0134221.
PR	21-OCT-1999;	9905-0160814.	PR	14-MAY-1999;	9905-0134370.
PR	21-OCT-1999;	9905-0160815.	PR	18-MAY-1999;	9905-0134768.
PR	22-OCT-1999;	9905-0160980.	PR	19-MAY-1999;	9905-0134941.
PR	22-OCT-1999;	9905-0160981.	PR	20-MAY-1999;	9905-0135124.
PR	22-OCT-1999;	9905-0160989.	PR	21-MAY-1999;	9905-0135353.
PR	25-OCT-1999;	9905-0161404.	PR	21-MAY-1999;	9905-0135629.
PR	25-OCT-1999;	9905-0161405.	PR	25-MAY-1999;	9905-0136021.
PR	25-OCT-1999;	9905-0161406.	PR	27-MAY-1999;	9905-0136392.
PR	26-OCT-1999;	9905-0161359.	PR	28-MAY-1999;	9905-0136782.
PR	26-OCT-1999;	9905-0161360.	PR	01-JUN-1999;	9905-0137222.
PR	26-OCT-1999;	9905-0161361.	PR	03-JUN-1999;	9905-0137528.
PR	28-OCT-1999;	9905-0161920.	PR	04-JUN-1999;	9905-0137502.
PR	28-OCT-1999;	9905-0161922.	PR	07-JUN-1999;	9905-0137724.
PR	28-OCT-1999;	9905-0161993.	PR	08-JUN-1999;	9905-0138094.
PR	28-OCT-1999;	9905-0162142.	PR	10-JUN-1999;	9905-0138540.
Query Match 45.1%; Score 41; DB 21; Length 1048;					9905-0138847.
Best Local Similarity 39.1%; Pred. No. 3.3e+02;					9905-0139119.
Matches 9; Conservative 2; Mismatches 4; Indels 8; Gaps 1;					9905-0139452.
QY	3 RRYSIGRYS-----VRFSWK 17		PR	16-JUN-1999;	9905-0139453.
Db	:		PR	17-JUN-1999;	9905-0139492.
	:		PR	18-JUN-1999;	9905-0139455.
	419 rkyssgkyqdhtgyrprvewk 441		PR	18-JUN-1999;	9905-0139456.
			PR	18-JUN-1999;	9905-0139457.
RESULT 14			PR	18-JUN-1999;	9905-0139458.
AAG42153			PR	18-JUN-1999;	9905-0139459.
ID	AAG42153 standard; Protein; 1129 AA.		PR	18-JUN-1999;	9905-0139460.
XX	AC	AAG42153;	PR	18-JUN-1999;	9905-0139461.
XX	AC		PR	18-JUN-1999;	9905-0139462.
XX	DT		PR	18-JUN-1999;	9905-0139463.
XX	DE	18-OCT-2000 (first entry)	PR	18-JUN-1999;	9905-0139750.
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 52534.	PR	18-JUN-1999;	9905-0139763.
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;	PR	21-JUN-1999;	9905-0139817.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		PR	22-JUN-1999;	9905-0139899.
KW	termination sequence.		PR	23-JUN-1999;	9905-0140353.
XX	OS		PR	23-JUN-1999;	9905-0140354.
XX	OS	Arabidopsis thaliana.	PR	24-JUN-1999;	9905-0140695.
PN	EP1033405-A2.		PR	28-JUN-1999;	9905-0140823.
PN			PR	29-JUN-1999;	9905-0140991.
PD	06-SEP-2000.		PR	30-JUN-1999;	9905-0141287.
XX	PF		PR	01-JUL-1999;	9905-0141842.
XX	PF	25-FEB-2000; 2000EP-0301439.	PR	01-JUL-1999;	9905-0142154.
XX	PR		PR	02-JUL-1999;	9905-0142055.
XX	PR	25-FEB-1999; 9905-0121825.	PR	06-JUL-1999;	9905-0142390.
PR	05-MAR-1999;	9905-0123180.	PR	08-JUL-1999;	9905-0142803.
PR	09-MAR-1999;	9905-0123548.	PR	12-JUL-1999;	9905-0142977.
PR	23-MAR-1999;	9905-0125788.	PR	13-JUL-1999;	9905-0143542.
PR	25-MAR-1999;	9905-0126264.	PR	14-JUL-1999;	9905-0143624.
PR	29-MAR-1999;	9905-0126785.	PR	15-JUL-1999;	9905-0144005.
PR	01-APR-1999;	9905-0127462.	PR	16-JUL-1999;	9905-0144085.
PR	06-APR-1999;	9905-0128234.	PR	16-JUL-1999;	9905-0144086.
PR	08-APR-1999;	9905-0128714.	PR	19-JUL-1999;	9905-0144325.
PR	16-APR-1999;	9905-0129845.	PR	19-JUL-1999;	9905-0144331.
PR	19-APR-1999;	9905-0130077.	PR	19-JUL-1999;	9905-0144332.
PR	21-APR-1999;	9905-0130449.	PR	19-JUL-1999;	9905-0144333.
PR	23-APR-1999;	9905-0130510.	PR	19-JUL-1999;	9905-0144334.
PR	23-APR-1999;	9905-0130891.	PR	20-JUL-1999;	9905-0144335.
PR	28-APR-1999;	9905-0131449.	PR	20-JUL-1999;	9905-0144632.
PR	30-APR-1999;	9905-0132048.	PR	20-JUL-1999;	9905-0144884.
PR	30-APR-1999;	9905-0132407.	PR	21-JUL-1999;	9905-0144814.
PR	04-MAY-1999;	9905-0132484.	PR	21-JUL-1999;	9905-0145086.
PR	05-MAY-1999;	9905-0132485.	PR	21-JUL-1999;	9905-0145088.
PR	06-MAY-1999;	9905-0132486.	PR	22-JUL-1999;	9905-0145085.
PR	06-MAY-1999;	9905-0132487.	PR	22-JUL-1999;	9905-0145087.
			PR	22-JUL-1999;	9905-0145089.
			PR	22-JUL-1999;	9905-0145192.

```

PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 17-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 07-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155458.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0158923.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 45.1%; Score 41; DB 21; Length 1129;
Best Local Similarity 39.1%; Pred. No. 3.6e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 8; Gaps 1;

QY 3 RRYSGRYSS-----VRFSSWK 17
|:| | | | | | | |
Db 500 rkyssgkyqdhtgyrvrviewk 522

RESULT 15
AAG42152
ID AAG42152 standard; Protein; 1193 AA.
XX
AC AAG42152;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52533.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
```

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140895.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161320.  
PR 28-OCT-1999; 99US-0161392.  
PR 28-OCT-1999; 99US-0161393.  
PR 29-OCT-1999; 99US-0162142.

Query Match 45.1%; Score 41; DB 21; Length 1193;  
Best Local Similarity 39.1%; Pred. NO. 3.8e+02;  
Matches 9; Conservative 2; Mismatches 4; Indels 8; Gaps 1;

QY 3 RYISIGRYS-----VRESWK 17  
|:|:| |:| |:  
DB 564 rkyssgkyqdhptgyrprvewk 586

Search completed: February 12, 2002, 12:30:33  
Job time: 366 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:24 ; Search time 106.12 Seconds  
(without alignments)  
3.605 Million cell updates/sec

Title: US-09-485-571-30

Perfect score: 91

Sequence: 1 RRRYSIGRYSVRFWSK 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	48.4	19	4	US-09-128-344A-82
2	44	48.4	19	4	US-09-128-344A-83
3	44	48.4	19	4	US-09-128-344A-181
4	44	48.4	19	4	US-09-128-344A-182
5	40.5	44.5	90	4	US-08-679-493A-93
6	40.5	44.5	101	4	US-08-679-493A-26
7	40	44.0	18	4	US-09-128-344A-79
8	40	44.0	18	4	US-09-128-344A-85
9	40	44.0	18	4	US-09-128-344A-177
10	40	44.0	18	4	US-09-128-344A-184
11	40	44.0	427	4	US-09-521-668B-18
12	40	44.0	427	4	US-09-521-668B-20
13	39	42.9	154	4	US-09-193-104-21
14	38	41.8	14	4	US-09-461-697-391
15	38	41.8	19	4	US-09-128-344A-81
16	38	41.8	19	4	US-09-128-344A-180
17	38	41.8	125	1	US-08-473-981A-10
18	38	41.8	125	2	US-08-474-087-10
19	38	41.8	254	4	US-09-240-915-7
20	38	41.8	254	4	US-09-591-435-7
21	38	41.8	568	4	US-09-188-930-291
22	38	41.8	900	4	US-08-890-865A-4
23	38	41.8	944	2	US-08-867-941-24
24	38	41.8	944	4	US-09-074-658-24
25	38	41.8	1487	2	US-08-760-489-2
26	38	41.8	1487	2	US-08-760-489-4
27	37.5	41.2	154	4	US-09-193-104-26

28	37.5	41.2	647	1	US-07-894-212A-8	Sequence 8, Appl
29	37.5	41.2	649	1	US-07-894-212A-2	Sequence 2, Appl
30	37.5	41.2	650	1	US-07-893-928A-1	Sequence 1, Appl
31	37	40.7	510	5	PCT-US96-03916-3	Sequence 3, Appl
32	37	40.7	623	5	PCT-US96-03916-64	Sequence 64, Appl
33	37	40.7	855	4	US-08-890-865A-10	Sequence 10, Appl
34	36.5	40.1	528	2	US-08-793-229-35	Sequence 35, Appl
35	36.5	40.1	528	3	US-09-285-957-35	Sequence 35, Appl
36	36	39.6	16	2	US-08-685-589A-207	Sequence 207, App
37	36	39.6	165	1	US-08-051-142-2	Sequence 2, Appl
38	36	39.6	410	4	US-08-725-758A-4	Sequence 4, Appl
39	36	39.6	426	4	US-08-725-758A-2	Sequence 2, Appl
40	35.5	39.0	336	1	US-08-289-699A-5	Sequence 5, Appl
41	35.5	39.0	336	2	US-08-878-283-5	Sequence 5, Appl
42	35	38.5	37	1	US-08-363-311-17	Sequence 17, Appl
43	35	38.5	37	2	US-08-463-288A-17	Sequence 17, Appl
44	35	38.5	37	2	US-08-470-445A-17	Sequence 17, Appl
45	35	38.5	37	5	PCT-US93-10506-17	Sequence 17, Appl

ALIGNMENTS

RESULT  
US-09-128-344A-82  
; Sequence 82, Application US/09128344A  
; Patent No. 6307016  
; GENERAL INFORMATION:  
; APPLICANT: Lehrer, Robert I.  
; APPLICANT: Harwig, Sylvia S. L.  
; APPLICANT: Chang, Conway C.  
; APPLICANT: Gu, Chee L.  
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS  
; CURRENT APPLICATION NUMBER: US/09/128.344A  
; CURRENT FILING DATE: 1998-08-03  
; PRIOR APPLICATION NUMBER: US 08/647,622  
; PRIOR FILING DATE: 1996-07-03  
; PRIOR APPLICATION NUMBER: US 60/000,898  
; PRIOR FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 82  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic polypeptide  
US-09-128-344A-82

Query Match 48.4%; Score 44; DB 4; Length 19;  
Best Local Similarity 63.6%; Pred. No. 0.56;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Oy 4 RYSIGRYSVRF 14  
Db 6 RYKGRFKVRF 16

RESULT  
US-09-128-344A-83  
; Sequence 83, Application US/09128344A  
; Patent No. 6307016  
; GENERAL INFORMATION:  
; APPLICANT: Lehrer, Robert I.  
; APPLICANT: Harwig, Sylvia S. L.  
; APPLICANT: Chang, Conway C.  
; APPLICANT: Gu, Chee L.  
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS  
; FILE REFERENCE: 8067-0053-999  
; CURRENT APPLICATION NUMBER: US/09/128.344A  
; CURRENT FILING DATE: 1998-08-03

us-09-485-571-30.rai

Wed Feb 13 07:52:30 2002

```

; APPLICANT: Chang, Conway C.
; APPLICANT: Gu, Chee L.
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS
; FILE REFERENCE: 8067-0053-999
; CURRENT APPLICATION NUMBER: US/09/128,344A
; CURRENT FILING DATE: 1998-08-03
; PRIOR APPLICATION NUMBER: US 08/647,622
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/000,898
; PRIOR FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic polypeptide
; NAME/KEY: MOD_RES
; LOCATION: 8,13
; OTHER INFORMATION: Xaa = MeGly
; US-09-128-344A-83

Query Match 48.4%; Score 44; DB 4; Length 19;
Best Local Similarity 63.6%; Pred. No. 0.56;
Matches 7; Conservative 2; Mismatches 2; Indels 2; Gaps 0;

QY 4 RYSIGRYSVRF 14
   ||:|:|
Db 6 RYXVGREXVRF 16

RESULT 3
US-09-128-344A-181
; Sequence 181, Application US/09128344A
; Patent No. 6307016
; GENERAL INFORMATION:
; APPLICANT: Lehrer, Robert I.
; APPLICANT: Harwig, Sylvia S. L.
; APPLICANT: Chang, Conway C.
; APPLICANT: Gu, Chee L.
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS
; FILE REFERENCE: 8067-0053-999
; CURRENT APPLICATION NUMBER: US/09/128,344A
; CURRENT FILING DATE: 1998-08-03
; PRIOR APPLICATION NUMBER: US 08/647,622
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/000,898
; PRIOR FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic polypeptide
; NAME/KEY: DISULFID
; LOCATION: (4)...(17)
US-09-128-344A-181

```

```

Query Match 48.4%; Score 44; DB 4; Length 19;
Best Local Similarity 63.6%; Pred. No. 0.56;
Matches 7; Conservative 2; Mismatches 2; Indels 2; Gaps 0;

QY 4 RYSIGRYSVRF 14
   ||:|:|
Db 6 RYXVGREXVRF 16

RESULT 5
US-08-679-493A-93
; Sequence 93, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; CURRENT FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-08-679-493A-93

```

```

Query Match 44.5%; Score 40.5; DB 4; Length 90;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 9; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 RSRYSI-GRYSVRF 17
   :|:|:|:|:|:|:|:|
Db 59 QSYRISGTYTCNNWK 76

RESULT 6
US-08-679-493A-26
; Sequence 26, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.

```

```

Query Match 48.4%; Score 44; DB 4; Length 19;
Best Local Similarity 63.6%; Pred. No. 0.56;
Matches 7; Conservative 2; Mismatches 2; Indels 2; Gaps 0;

QY 4 RYSIGRYSVRF 14
   ||:|:|
Db 6 RYXVGREXVRF 16

RESULT 4
US-09-128-344A-182
; Sequence 182, Application US/09128344A
; Patent No. 6307016
; GENERAL INFORMATION:
; APPLICANT: Lehrer, Robert I.
; APPLICANT: Harwig, Sylvia S. L.

```

## ; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS

; FILE REFERENCE: 55-95  
; CURRENT APPLICATION NUMBER: US/08/679,493A  
; CURRENT FILING DATE: 1996-07-12  
; PRIOR APPLICATION NUMBER: 60/001203  
; PRIOR FILING DATE: 1995-07-14  
; PRIOR APPLICATION NUMBER: 60/003,112  
; PRIOR FILING DATE: 1995-09-01  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 101  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)..(101)  
; OTHER INFORMATION: X is selenocysteine.  
US-08-679-493A-26

Query Match 44.5%; Score 40.5; DB 4; Length 101;  
Best Local Similarity 52.9%; Pred. No. 11;  
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Qy 2 SRRYSIGR-YSVRFSWK 17  
| |||| | : : ||  
Db 71 SYRYSISRTVTCQHNWK 87

## RESULT 7

US-09-128-344A-79  
; Sequence 79, Application US/09128344A  
; Patent No. 6307016  
; GENERAL INFORMATION:  
; APPLICANT: Lehner, Robert I.  
; APPLICANT: Harwig, Sylvia S. L.  
; APPLICANT: Chang, Conway C.  
; APPLICANT: Gu, Chee L.  
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS  
; FILE REFERENCE: 8067-0053-999  
; CURRENT APPLICATION NUMBER: US/09/128,344A  
; CURRENT FILING DATE: 1998-08-03  
; PRIOR APPLICATION NUMBER: US 08/647,622  
; PRIOR FILING DATE: 1996-07-03  
; PRIOR APPLICATION NUMBER: US 60/000,898  
; PRIOR FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 79  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic polypeptide  
US-09-128-344A-79

Query Match 44.0%; Score 40; DB 4; Length 18;  
Best Local Similarity 60.0%; Pred. No. 2.3;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RYSIGRYSVR 13  
| : : | : : ||  
Db 6 RYAVGREAVR 15

## RESULT 8

US-09-128-344A-85  
; Sequence 85, Application US/09128344A  
; Patent No. 6307016  
; GENERAL INFORMATION:  
; APPLICANT: Lehner, Robert I.

; APPLICANT: Harwig, Sylvia S. L.  
; APPLICANT: Chang, Conway C.  
; APPLICANT: Gu, Chee L.  
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS  
; FILE REFERENCE: 8067-0053-999  
; CURRENT APPLICATION NUMBER: US/09/128,344A  
; CURRENT FILING DATE: 1998-08-03  
; PRIOR APPLICATION NUMBER: US 08/647,622  
; PRIOR FILING DATE: 1996-07-03  
; PRIOR APPLICATION NUMBER: US 60/000,898  
; PRIOR FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 85  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic polypeptide  
US-09-128-344A-85

Query Match 44.0%; Score 40; DB 4; Length 18;  
Best Local Similarity 60.0%; Pred. No. 2.3;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RYSIGRYSVR 13  
| : : | : : ||  
Db 6 RYAVGREAVR 15

## RESULT 9

US-09-128-344A-177  
; Sequence 177, Application US/09128344A  
; Patent No. 6307016  
; GENERAL INFORMATION:  
; APPLICANT: Lehner, Robert I.  
; APPLICANT: Harwig, Sylvia S. L.  
; APPLICANT: Chang, Conway C.  
; APPLICANT: Gu, Chee L.  
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS  
; FILE REFERENCE: 8067-0053-999  
; CURRENT APPLICATION NUMBER: US/09/128,344A  
; CURRENT FILING DATE: 1998-08-03  
; PRIOR APPLICATION NUMBER: US 08/647,622  
; PRIOR FILING DATE: 1996-07-03  
; PRIOR APPLICATION NUMBER: US 60/000,898  
; PRIOR FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 177  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic polypeptide  
; NAME/KEY: DISULFID  
; LOCATION: (4)..(17)  
US-09-128-344A-177

Query Match 44.0%; Score 40; DB 4; Length 18;  
Best Local Similarity 60.0%; Pred. No. 2.3;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RYSIGRYSVR 13  
| : : | : : ||  
Db 6 RYAVGREAVR 15

## RESULT 10

US-09-128-344A-184  
; Sequence 184, Application US/09128344A

```
; Patent No. 6307016
; GENERAL INFORMATION:
; APPLICANT: Lehrer, Robert I.
; APPLICANT: Harwig, Sylvia S. L.
; APPLICANT: Chang, Conway C.
; APPLICANT: Gu, Chee L.
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS
; FILE REFERENCE: 8067-0053-999
; CURRENT APPLICATION NUMBER: US/09/128,344A
; CURRENT FILING DATE: 1998-08-03
; PRIOR APPLICATION NUMBER: US 08/647,622
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/000,898
; PRIOR FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 184
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic polypeptide
; NAME/KEY: DISULFID
; LOCATION: (4)...(17)
US-09-128-344A-184

Query Match 44.0%; Score 40; DB 4; Length 18;
Best Local Similarity 60.0%; Pred. No. 2.3;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RYISGRYSVR 13
Db 6 RYAVGRFAVR 15

RESULT 11
US-09-521-668B-18
; Sequence 18, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: KANNO, SOHEI
; APPLICANT: KIMURA, EIICHIRO
; APPLICANT: MATSUI, KAHIKO
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
; FILE REFERENCE: 0010-1093-0
; CURRENT APPLICATION NUMBER: US/09/521,668B
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: JP 11-69896
; PRIOR FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
US-09-521-668B-18

Query Match 44.0%; Score 40; DB 4; Length 427;
Best Local Similarity 43.8%; Pred. No. 61;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RSRYSIGRYSVRFSW 16
Db 276 RWREYEVGSKNLRSSW 291

RESULT 13
US-09-193-104-21
; Sequence 21, Application US/09193104A
; Patent No. 6172193
; GENERAL INFORMATION:
; APPLICANT: Primi, Daniele
; APPLICANT: Fiordalisi, Gianfranco
; APPLICANT: Palla, Mario
; TITLE OF INVENTION: Escape Mutant of the Surface Antigen of Hepatitis B
; FILE REFERENCE: SEDI004US
; CURRENT APPLICATION NUMBER: US/09/193,104A
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: EP 97830635.5
; EARLIER FILING DATE: 1997-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Hepatitis B virus
US-09-193-104-21

Query Match 42.9%; Score 39; DB 4; Length 154;
Best Local Similarity 61.5%; Pred. No. 30;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 4 RYISGRYSVRFSW 16
Db 127 KYLWERSVRFSW 139

RESULT 14
US-09-461-697-391
; Sequence 391, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
```



; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH

; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 391  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-461-697-391

Query Match 41.8%; Score 38; DB 4; Length 14;  
Best Local Similarity 70.0%; Pred. No. 3.6;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 YSIGRYSVRF 14  
|||:||||  
Db 5 YSGKYSIRF 14

RESULT 15  
US-09-128-344A-81  
; Sequence 81, Application US/09128344A  
; Patent No. 6307016  
; GENERAL INFORMATION:  
; APPLICANT: Lehrer, Robert I.  
; APPLICANT: Harwig, Sylvia S. L.  
; APPLICANT: Chang, Conway C.  
; APPLICANT: Gu, Chee L.  
; TITLE OF INVENTION: PAREVINS AND TACHYTEGRINS  
; FILE REFERENCE: 8067-0053-999  
; CURRENT APPLICATION NUMBER: US/09/128,344A  
; CURRENT FILING DATE: 1998-08-03  
; PRIOR APPLICATION NUMBER: US 08/647,622  
; PRIOR FILING DATE: 1996-07-03  
; PRIOR APPLICATION NUMBER: US 60/000,898  
; PRIOR FILING DATE: 1995-07-06  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 81  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic polypeptide  
US-09-128-344A-81

Query Match 41.8%; Score 38; DB 4; Length 19;  
Best Local Similarity 54.5%; Pred. No. 4.9;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 RYSGRYSVRF 14  
|||:||||  
Db 6 RYTVPRETVRF 16

Search completed: February 12, 2002, 12:32:24  
Job time: 452 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:41 ; Search time 126.85 Seconds  
(without alignments)  
10.209 Million cell updates/sec

Title: US-09-485-571-30

Perfect score: 91

Sequence: 1 RRRYSIGRYSVRFWSK 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68:\*

1: piri:\*

2: piri:\*

3: piri:\*

4: piri:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	48.4	124	2 T45688	hypothetical prote
2	43	47.3	130	2 I46641	rearranged T-cell
3	43	47.3	1364	2 T51920	probable xanthine
4	42	46.2	540	2 T48619	tRNA synthase-like
5	41	45.1	1193	2 A86193	hypothetical prote
6	41	45.1	1236	1 VHWWE	structural polypro
7	40	44.0	240	2 I40335	hypothetical prote
8	40	44.0	267	2 A71871	hypothetical prote
9	40	44.0	331	2 T22968	hypothetical prote
10	40	44.0	723	2 T47613	ABC transporter-li
11	40	44.0	855	2 F45557	external glycoprot
12	39.5	43.4	197	2 T40144	60s ribosomal prot
13	39	42.9	171	2 E84472	hypothetical prote
14	39	42.9	191	2 A71060	hypothetical prote
15	39	42.9	261	2 F85072	hypothetical prote
16	39	42.9	353	2 T25678	hypothetical prote
17	39	42.9	1134	2 D49851	magnesium-protopor
18	38.5	42.3	662	2 S36792	hypothetical prote
19	38	41.8	102	2 S09828	hypothetical prote
20	38	41.8	275	2 S03967	intercellular adhe
21	38	41.8	366	2 T24546	hypothetical prote
22	38	41.8	368	2 S19414	hypothetical prote
23	38	41.8	376	2 S70841	hypothetical trans
24	38	41.8	406	2 H75527	probable cation ex
25	38	41.8	491	2 H71097	hypothetical prote
26	38	41.8	581	2 H75196	hypothetical prote
27	38	41.8	643	2 I59412	seppA protein - Esc
28	38	41.8	675	2 E86046	escv [imported] -
29	38	41.8	708	2 T00205	hypothetical prote

30 38 41.8 832 2 H72278  
31 38 41.8 888 2 E65057  
32 38 41.8 899 2 H84639  
33 38 41.8 958 2 A82583  
34 38 41.8 1151 2 T24541  
35 38 41.8 1193 2 T50729  
36 38 41.8 1636 2 S54520  
37 38 41.8 1740 2 T43215  
38 38 41.8 2049 2 T29227  
39 37.5 41.2 226 1 J01579  
40 37.5 41.2 226 1 J01580  
41 37.5 41.2 226 2 J02119  
42 37.5 41.2 226 2 J02120  
43 37.5 41.2 226 2 J02225  
44 37.5 41.2 226 2 J02121  
45 37.5 41.2 226 2 J02122

alpha-mannosidase-  
hypothetical prote  
hypothetical prote  
conserved hypotet  
hypothetical prote  
magnesium-protopor  
probable membrane  
ribonucleotide red  
hypothetical prote  
major surface anti  
major surface anti  
surface antigen -  
small surface prot  
surface antigen -  
surface antigen -

ALIGNMENTS

RESULT 1  
T45688  
hypothetical protein F18L15.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45688  
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.  
submitted to the Protein Sequence Database, November 1999  
A:Reference number: Z23010  
A:Accession: T45688  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-124 <CHO>  
A:CROSS-references: EMBL:AL133298  
A:Experimental source: cultivar Columbia; BAC clone F18L15  
C:Genetics:  
A:Map position: 3  
A:Note: F13L15.30

Query March 48.4%; Score 44; DB 2; Length 124;  
Best Local Similarity 50.0%; Pred. No. 3.2;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 SRRYSIGRYSVRFWSK 17  
Db 95 SKRASLKGFFVRFTRK 110  
1:1 | : : |||:|  
2 SKRASLKGFFVRFTRK 110

RESULT 2  
I46641  
rearranged T-cell receptor delta-chain/ vdelta4-Ddeltas-delta4 - pig (fragment)  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 23-Jul-1999  
C:Accession: I46641  
R:Yang, Y.G.; Ohta, S.; Yamada, S.; Shimizu, M.; Takagaki, Y.  
J. Immunol. 155, 1981-1993, 1995  
A:Title: Diversity of T cell receptor delta-chain cDNA in the thymus of a one-month  
A:Reference number: I46623; MUID:95363165  
A:Accession: I46641  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-130 <YAN>  
A:CROSS-references: GB:D49582; NID:g1041166; PIDN:BA08526.1; PID:g1041167  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: T-cell receptor

Query Match 47.3%; Score 43; DB 2; Length 130;  
Best Local Similarity 64.3%; Pred. No. 5;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RSRYSIGRYSVRF 14  
 Db 63 QNRNREGYSVEF 76

RESULT 3  
 T51920  
 probable xanthine dehydrogenase [imported] - Neurospora crassa  
 N:Alternate names: protein B23111.320  
 C:Species: Neurospora crassa  
 C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 01-Dec-2000  
 C:Accession: T51920  
 R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, August 2000  
 A:Reference number: Z25858  
 A:Accession: T51920  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1364 <SCH>  
 A:Cross-references: EMBL:AL391572; GSPDB:GN00116; NCSP:B23111.320  
 A:Experimental source: BAC clone B23111, strain OR74A  
 C:Genetics:  
 A:Gene: NCSP:B23111.320  
 A:Map position: 6  
 A:Introns: 66/2, 1321/3  
 C:Superfamily: xanthine dehydrogenase; ferredoxin (2Fe-2S) homology  
 C:Keywords: 2Fe-2S; metalloprotein  
 F:68,73,76,98/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 47.3%; Score 43; DB 2; Length 1364;  
 Best Local Similarity 43.8%; Pred. No. 53;  
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRF 17  
 Db 1007 SRRALANYVEHKWR 1022

RESULT 4  
 T48619  
 tRNA synthase-like protein - Arabidopsis thaliana  
 N:Alternate names: protein F18022.250  
 C:Species: Arabidopsis thaliana (mouse-ear cross)  
 C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T48619  
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z24493  
 A:Accession: T48619  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-540 <BEV>  
 A:Cross-references: EMBL:DB  
 A:Experimental source: cultivar Columbia; BAC clone F18022  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 331/1; 349/3; 377/2; 403/3; 440/3; 477/2; 505/2; 518/1  
 A:Note: F18022.250

Query Match 46.2%; Score 42; DB 2; Length 540;  
 Best Local Similarity 53.3%; Pred. No. 30;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 RRYSIGRYSVRF 17  
 Db 513 RRDSTGEYSANDAW 527

RESULT 5  
 A86193  
 hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: A86193  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alic Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kiu C.A.; Li, J.H.; Li, V.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzl Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tail ket, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: A86193  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1193 <STO>  
 A:Cross-references: GB:AE005172; NID:g6850313; PIDN:AAF29390.1; GSPDB:GN00141/  
 C:Genetics:  
 A:Map position: 1

Query Match 45.1%; Score 41; DB 2; Length 1193;  
 Best Local Similarity 39.1%; Pred. No. 99;  
 Matches 9; Conservative 2; Mismatches 4; Indels 8; Gaps 1;

QY 3 RRYSIGRYS- - - - - VRFWSK 17  
 Db 564 RYSSGKYDHPGTGYRVRVEWK 586

RESULT 6  
 VHWWE  
 structural polyprotein - western equine encephalomyelitis virus (strain BFS1703)  
 N:Contains: 6K protein; coat protein C; membrane glycoprotein E1; membrane glycopro C:Species: western equine encephalomyelitis virus  
 C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 16-Jul-1999  
 C:Accession: A35587  
 R:Hahn, C.S.; Lustig, S.; Strauss, E.G.; Strauss, J.H.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 5997-6001, 1988  
 A:Title: Western equine encephalitis virus is a recombinant virus.  
 A:Reference number: A35587; MUID:88320369  
 A:Accession: A35587  
 A:Molecule type: genomic RNA  
 A:Residues: 1-1236 <HAH>  
 A:Cross-references: GB:J03854; NID:g323728; PIDN:AAA42999.1; PID:g323730  
 C:Superfamily: togavirus structural polyprotein  
 C:Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein  
 F:1-259/Product: coat protein C #status predicted <CPC>  
 F:260-319/Product: membrane glycoprotein E1 #status predicted <MG3>  
 F:320-742/Product: membrane glycoprotein E2 #status predicted <MG2>  
 F:684-702/Product: transmembrane #status predicted <TN1>  
 F:719-737/Product: transmembrane #status predicted <TN2>  
 F:743-797/Product: 6K protein #status predicted <6KP>  
 F:758-775/Product: transmembrane #status predicted <TN3>  
 F:777-793/Product: transmembrane #status predicted <TN4>  
 F:798-1236/Product: membrane glycoprotein E1 #status predicted <MG1>  
 F:1206-1227/Product: transmembrane #status predicted <TN5>  
 F:50,270,515,637,724,936,1042/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 45.1%; Score 41; DB 1; Length 1236;  
 Best Local Similarity 40.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRF 16  
 Db 635 SRNFSVGRGLEYVW 649

RESULT 7  
 I40335

hypothetical protein 1 - Corynebacterium glutamicum  
 C:Species: Corynebacterium glutamicum  
 C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 22-Oct-1999  
 C:Accession: I40335  
 R:Yamaguchi, R.; Terabe, M.; Miwa, K.; Tsuchiya, M.; Takagi, H.; Morinaga, Y.; Nakamori, A.; Agric. Biol. Chem. 50, 2771-2778, 1986  
 A:Title: Determination of the complete nucleotide sequence of Brevibacterium lactofermentum  
 A:Reference number: I40334  
 A:Accession: I40335  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-240 <RES>  
 A:Cross-references: GB:D00038; NID:g216381; PIDN:BAA00026.1; PID:g216383

Query Match 44.0%; Score 40; DB 2; Length 240;  
 Best Local Similarity 43.8%; Pred. No. 29;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 RSRYSIGRYSVRFWSW 16  
 I I I I I : I I I I  
 Db 89 RWREYEVGSKNLRSSW 104

RESULT 8  
 A71871  
 hypothetical protein jhp0933 - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
 C:Accession: A71871  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen  
 A:Reference number: A71800; MUID:99120557  
 A:Accession: A71871  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-267 <ARN>  
 A:Cross-references: GB:AE001522; GB:AE001439; NID:g4155505; PIDN:AAD06511.1; PID:g4155511  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: jhp0933

Query Match 44.0%; Score 40; DB 2; Length 267;  
 Best Local Similarity 58.3%; Pred. No. 32;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 SIGRYSVRFWSW 17  
 I I I I I I I I I I  
 Db 91 SVGRYVRYATK 102

RESULT 9  
 T22968  
 hypothetical protein F59A1.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T22968  
 R:Mortimore, B.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19644  
 A:Accession: T22968  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-331 <WIL>  
 A:Cross-references: EMBL:Z81557; PIDN:CAB04537.1; GSPDB:GN00023; CESP:F59A1.9  
 A:Experimental source: clone F59A1  
 C:Genetics:  
 A:Gene: CESP:F59A1.9  
 A:Map position: 5

A:Introns: 69/2; 97/2; 242/3

Query Match 44.0%; Score 40; DB 2; Length 331;  
 Best Local Similarity 50.0%; Pred. No. 40;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 RYSIGRYSVRFWSW 17  
 : I I I I I I I I  
 Db 122 KFDIGEYSLRTSPK 135

RESULT 10  
 T47613  
 ABC transporter-like protein - Arabidopsis thaliana  
 N:Alternate names: protein T14E10.110  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T47613  
 R:Obermaier, B.; Ottenwaeider, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Lemcke, submitted to the Protein Sequence Database, March 2000  
 A:Reference number: Z24463  
 A:Accession: T47613  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-723 <OBE>  
 A:Cross-references: EMBL:AL138656  
 A:Experimental source: cultivar Columbia; BAC clone T14E10  
 C:Genetics:  
 A:Map position: 3  
 A:Note: T14E10.110

Query Match 44.0%; Score 40; DB 2; Length 723;  
 Best Local Similarity 57.1%; Pred. No. 88;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RSRYSIGRYSVRF 14  
 I I I I I I I I I I  
 Db 556 FSQKLRIGRYSOHF 569

RESULT 11  
 F45557  
 external glycoprotein, transmembrane protein(external glycoprotein, EGP, transmembrane glycoprotein, transmembrane protein)  
 C:Species: feline immunodeficiency virus  
 C:Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Feb-1997  
 C:Accession: F45557  
 R:Maki, N.; Miyazawa, T.; Fukasawa, M.; Hasegawa, A.; Hayami, M.; Miki, K.; Mikami, A.; Arch. Virol. 123, 29-45, 1992  
 A:Title: Molecular characterization and heterogeneity of feline immunodeficiency virus  
 A:Reference number: A45557; MUID:92198230  
 A:Accession: F45557  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-855 <NAK>  
 A:Experimental source: strain TM2  
 A:Note: sequence extracted from NCBI backbone (NCBI:89826, NCBI:89854)  
 C:Superfamily: feline immunodeficiency virus env polyprotein  
 C:Keywords: glycoprotein; transmembrane protein

Query Match 44.0%; Score 40; DB 2; Length 855;  
 Best Local Similarity 37.5%; Pred. No. 1e+02;  
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSRYSIGRYSVRFWSW 16  
 I I I I I I I I I I  
 Db 449 KSKRHSERFRICKW 464

RESULT 12  
 T40144

Wed Feb 13 07:52:31 2002

A:Molecule type: DNA  
A:Residues: 1-191 <XAW>  
A:Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30275.1; PID:dl031218; PID:93  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C:Species: Schizosaccharomyces pombe  
C:Accession: T40144  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21842  
A:Accession: T40144  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <WOO>  
A:Cross-references: EMBL:AL022103; PIDN:CAA17885.1; GSPDB:GN00067; SPDB:SPBC2G2.05  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C:Species: Pyrococcus horikoshii  
C:Accession: A71060  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71060  
A:Accession: A71060  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

60s ribosomal protein l16-c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C:Accession: T40144  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21842  
A:Accession: T40144  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <WOO>  
A:Cross-references: EMBL:AL022103; PIDN:CAA17885.1; GSPDB:GN00067; SPDB:SPBC2G2.05  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C:Species: Pyrococcus horikoshii  
C:Accession: A71060  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71060  
A:Accession: A71060  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

Query Match 43.4%; Score 39.5; DB 2; Length 197;  
Best Local Similarity 52.9%; Pred. No. 29;  
Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;  
QY 2 SRRY-SIGRYSVRFWSK 17  
||| :||| | ||  
Db 133 SRKYCTIGRLSSEVGWK 149

RESULT 13  
E84472  
hypothetical protein At2g05880 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84472  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: E84472  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-171 <STO>  
A:Cross-references: GB:AE002093; NID:g4006823; PIDN:AAC95165.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g05880  
A:Map position: 2

Query Match 42.9%; Score 39; DB 2; Length 171;  
Best Local Similarity 42.9%; Pred. No. 30;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 4 RYSIGRYSVRFWSK 17  
||| | :||| |  
Db 102 RYSSFRFHIFWE 115

RESULT 14  
A71060  
hypothetical protein PH1175 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C:Accession: A71060  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71060  
A:Accession: A71060  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

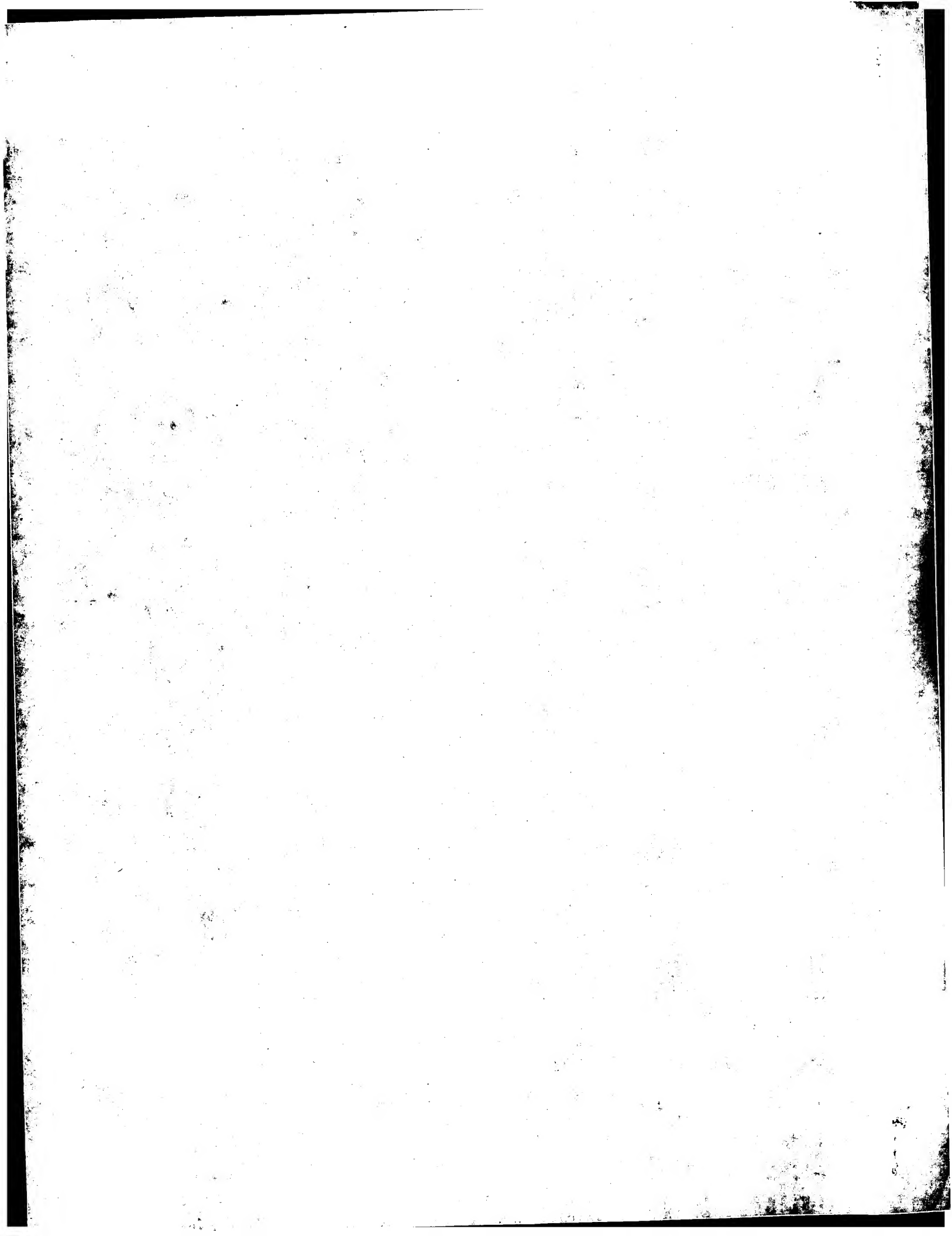
Query Match 42.9%; Score 39; DB 2; Length 191;  
Best Local Similarity 50.0%; Pred. No. 34;  
Matches 9; Conservative 1; Mismatches 4; Indels 4; Gaps 1;  
QY 3 RRY-----SIGRYSVRFWSW 16  
||| | :||| |  
Db 134 RRYFYSSWSKRVPRFEW 151

RESULT 15  
F85072  
hypothetical protein AT4g07440 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: F85072  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488.  
A:Accession: F85072  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267340; PIDN:CAB81114.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g07440  
A:Map position: 4

Query Match 42.9%; Score 39; DB 2; Length 261;  
Best Local Similarity 42.9%; Pred. No. 46;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 4 RYSIGRYSVRFWSK 17  
||| | :||| |  
Db 87 RYSSFRFHIFWE 100

Search completed: February 12, 2002, 12:34:42  
Job time: 560 sec

---







Wed Feb 13 07:52:32 2002

us-09-485-571-30.rsp

```

FT CARBOHYD 565 565 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 794 AA; 87349 MW; A4E84E17896C168D CRC64;

Query Match 48.4%; Score 44; DB 1; Length 794;
Best Local Similarity 75.0%; Pred. No. 6.5;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RSRYSIGRYSV 12
    ||| |::|||
Db 714 RSRAASVGRYSV 725

RESULT 2
POLYS_WEEV STANDARD; PRT: 1236 AA.
AC AC POLYS_WEEV 088696; Q88697; Q88698; Q88699; Q88700;
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE STRUCTURAL POLYPEPTIDE (P130) [CONTAINS: COAT PROTEIN C (EC 3.4.21.-)
DE [CAPSID PROTEIN C); SPIKE GLYCOPROTEIN E3; SPIKE GLYCOPROTEIN E2;
DE 6 KDA PEPTIDE; SPIKE GLYCOPROTEIN E1].
OS Western equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BFS1703;
RX MEDLINE=88320369; PubMed=3413072;
RA Hahn C.S., Lustig S., Strauss E.G., Strauss J.H.;
RA "Western equine encephalitis virus is a recombinant virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:5997-6001(1988).
CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC HEMAGGLUTININ.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J03854; AAA42999.1;
CC PIR; A35587; VHWWE.
CC HSP; P03316; 1WK.
CC
CC MEROPS; S03.001;
CC
CC InterPro: IPR002548; Alpha_E1_glycop.
CC InterPro: IPR000936; Alpha_E2_glycop.
CC InterPro: IPR002533; Alpha_E3_glycop.
CC InterPro: IPR001836; Alpha_Core.
CC InterPro: IPR000930; Togavirin.
CC Pfam; PF00944; Alpha_Core; 1.
CC Pfam; PF01589; Alpha_E1_glycop; 1.
CC Pfam; PF00943; Alpha_E2_glycop; 1.
CC Pfam; PF01563; Alpha_E3_glycop; 1.
CC
CC PRINTS; PR00798; TOGAVIRIN; Transmembrane; Glycoprotein; Hydrolase;
CC Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
CC Serine protease.
CC CHAIN 1 259 COAT PROTEIN C.
CC CHAIN 260 319 SPIKE GLYCOPROTEIN E3.
CC CHAIN 320 742 SPIKE GLYCOPROTEIN E2.
CC CHAIN 743 797 6 KDA PEPTIDE.
CC CHAIN 798 1236 SPIKE GLYCOPROTEIN E1.
CC CHAIN 136 136 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```

FT TRANSMEM 684 702 POTENTIAL.
FT TRANSMEM 719 737 POTENTIAL.
FT TRANSMEM 758 775 POTENTIAL.
FT TRANSMEM 777 793 POTENTIAL.
FT TRANSMEM 1206 1227 POTENTIAL.
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1042 1042 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1236 AA; 136082 MW; 0756DB0B0A1CCA96 CRC64;

Query Match 45.1%; Score 41; DB 1; Length 1236;
Best Local Similarity 40.0%; Pred. No. 34;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRFWS 16
    ||::|||::|
Db 635 SRNFSVGRGLETVW 649

RESULT 3
ENV_FIVT2 STANDARD; PRT: 855 AA.
ID ENV_FIVT2
AC Q02282;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPEPTIDE PRECURSOR (GP150 POLYPEPTIDE) [CONTAINS: GLYCOPROTEIN
DE GP100; GLYCOPROTEIN GP36].
GN ENV.
OS Feline immunodeficiency virus (isolate TM2) (FIV).
OC Viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=31676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91303718; PubMed=1649349;
RA Kiyonasu T., Miyazawa T., Furuya T., Shibata R., Sakai H.,
RA Sakurai J.I., Fukasawa M., Maki N., Hasegawa A., Mikami T.,
RA Adachi A.;
RA "Identification of feline immunodeficiency virus rev gene activity.";
RA J. Virol. 65:4539-4542(1991).
CC [2]
CC SEQUENCE FROM N.A.
CC RP MEDLINE=92198230; PubMed=1312825;
CC RA Maki N., Miyazawa T., Fukasawa M., Hasegawa A., Hayami M., Miki K.,
CC RA Mikami T.;
CC "Molecular characterization and heterogeneity of feline
CC immunodeficiency virus isolates.";
CC Arch. Virol. 123:29-45(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M59418; AAA43074.1;
CC PIR; F45557; F45557.
CC InterPro: IPR002050; Env_polyprotein.
CC Fram; PR00429; Env_polyprotein; 1.
CC Coat protein; Polyprotein; Glycoprotein; Transmembrane.
CC CHAIN 1 610 MAJOR GLYCOPROTEIN GP100.
CC CHAIN 611 855 GLYCOPROTEIN GP36.
CC TRANSMEM 786 809 POTENTIAL.
CC CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```
FT CARBOHYD 298 298 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 342 342 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 716 716 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 720 720 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 736 736 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLNAC. . .) (POTENTIAL).
SQ SEQUENCE 855 AA; 98238 MW; BC2DFABB6245D70D CRC64;

Query Match 44.0%; Score 40; DB 1; Length 855;
Best Local Similarity 37.5%; Pred. No. 34;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSRRYSGRYSVRESW 16
   :|:|:| |:|:| |
DB 449 KSRHSEARFIRCKW 464

RESULT 4
R16C_SCHPO STANDARD; PRT; 197 AA.
AC Q43004;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L16-C.
GN RPL16C OR SPBC262.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Devlin K., Churcher C.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- MISCELLANEOUS: THERE ARE THREE GENES FOR L16 IN S. POMBE.
CC -1- SIMILARITY: BELONGS TO THE L13P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL022103; CAA17885.1; -.
CC InterPro: IPR001074; Ribosomal_L13.
CC Pfam: PF00572; Ribosomal_L13; 1.
CC ProDom: PD001791; Ribosomal_L13; 1.
CC PROSITE: PS00783; RIBOSOMAL_L13; 1.
CC Ribosomal protein; Multigene family.
SQ SEQUENCE 197 AA; 22353 MW; C529CC6BB55E77D6 CRC64;

Query Match 43.4%; Score 39.5; DB 1; Length 197;
Best Local Similarity 52.9%; Pred. No. 8.9;

Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 RYSIGRYSVRESWK 17
   ||:|:|:| |:
DB 199 RYLGVYADNRDWR 212

RESULT 6
```

```
Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 2 SRRY-SIGRYSVRESWK 17
   ||:|:|:| |:
DB 133 SRRYCTIGRLSSEVGWK 149

RESULT 5
BCHL_RHOCA STANDARD; PRT; 1194 AA.
ID BCHL_RHOCA
AC P26132;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAGNESIUM-CHELATASE SUBUNIT H (MG-PROTOPORPHYRIN IX CHELATASE SUBUNIT
DE H).
DE BCHL.
GN Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OS Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003;
RX MEDLINE=93224465; PubMed=8385667;
RA Burke D.H., Alberti M., Hearst J.E.;
RT "bchfNBH bacteriochlorophyll synthesis genes of Rhodobacter
RT capsulatus and identification of the third subunit of
RT light-independent protochlorophyllide reductase in Bacteria and
RT plants.";
RT J. Bacteriol. 175:2414-2422(1993).
RN [2]
RP SEQUENCE OF 1170-1194 FROM N.A.
RX MEDLINE=90368552; PubMed=2203738;
RA Yang Z., Bauer C.E.;
RT "Rhodobacter capsulatus genes involved in early steps of the
RT bacteriochlorophyll biosynthetic pathway.";
RL J. Bacteriol. 172:5001-5010(1990).
CC -1- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES
CC A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-
CC PHOTOPORPHYRIN IX.
CC -1- PATHWAY: BACTERIOCHLOROPHYLL BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNIT H FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11165; CAA77524.1; -.
CC EMBL: M34843; AAA26097.1; -.
CC PIR: S17808; S17808.
CC PIR: A36716; A36716.
CC PIR: D49851; D49851.
CC InterPro: IPR003672; CbnM/Mg_chelatase.
CC Pfam: PF02514; cbnM-Mg_chel; 1.
CC Photosynthesis; Chlorophyll biosynthesis.
SQ SEQUENCE 1194 AA; 129361 MW; 6341816A58774EE5 CRC64;

Query Match 42.9%; Score 39; DB 1; Length 1194;
Best Local Similarity 42.9%; Pred. No. 72;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 RYSIGRYSVRESWK 17
   ||:|:|:| |:
DB 199 RYLGVYADNRDWR 212

RESULT 6
```

22

FT DOMAIN 22 223 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 224 248 POTENTIAL.  
 FT DOMAIN 249 275 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 41 98 IG-LIKE C2-TYPE DOMAIN 1.  
 FT DOMAIN 127 197 IG-LIKE C2-TYPE DOMAIN 2.  
 FT DISULFID 48 91  
 FT DISULFID 52 95  
 FT DISULFID 134 190  
 FT CARBOHYD 47 47  
 FT CARBOHYD 82 82  
 FT CARBOHYD 105 105  
 FT CARBOHYD 153 153  
 FT CARBOHYD 176 176  
 FT CARBOHYD 187 187  
 SQ SEQUENCE 275 AA; 30653 MW; 2F0F142732019152 CRC64;

Query Match 41.8%; Score 38; DB 1; Length 275;  
 Best Local Similarity 40.0%; Pred. No. 23;  
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 RYSGRYSVRFWSK 17  
 I: | | | | | | | | | |  
 Db 253 RQRMCTYGVRAWR 267

RESULT 9  
 YC22\_YEAST  
 AC P25608; STANDARD; PRT; 368 AA.  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 40.1 KDA PROTEIN IN GIT1-PAU3 INTERGENIC REGION.  
 GN YCR102C  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grivell L.A., de Haan M., Maat M.J.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE YCR102C/YLR460C/YNL134C FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X59720; CAA42244.1; -;  
 CC PIR; S19414; S19414.  
 CC YEPD; 5470; -;  
 CC SGD; S0000699; YCR102C.  
 CC InterPro; IPR002085; Adh\_zinc.  
 CC Pfam; PF00107; adh\_zinc; 1  
 SQ SEQUENCE 368 AA; 40121 MW; 0BEE4FB4DB04AF8B CRC64;

Query Match 41.8%; Score 38; DB 1; Length 368;  
 Best Local Similarity 57.1%; Pred. No. 31;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RSRRYSGRYSVRF 14  
 I: | | | | | | | | | |  
 Db 287 RTRLRYSGHGVPF 300

RESULT 10  
 YGCB\_ECOLI

ID YGCB\_ECOLI STANDARD; PRT; 888 AA.  
 AC P38036; Q45902;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL 100.5 KDA PROTEIN IN IAP-CYSH INTERGENIC REGION.  
 GN YGCB OR B2761.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau E., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE OF 1-335 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91172132; PubMed=2005873;  
 RA Krone F.A., Westphal G., Schwenk J.D.;  
 RT "Characterisation of the gene *cysH* and of its product phospho-  
 RT adenylylsulphate reductase from Escherichia coli.";  
 RL Mol. Gen. Genet. 225:314-319(1991).  
 RN [3]  
 RN IDENTIFICATION.  
 RP MEDLINE=95075659; PubMed=7984428;  
 RX Borodovsky M., Rudd K.E., Koonin E.V.;  
 RA "Intrinsic and extrinsic approaches for detecting genes in a  
 RT bacterial genome.";  
 RL Nucleic Acids Res. 22:4756-4767(1994).  
 CC -!- SIMILARITY: BELONGS TO THE PULE/OUTE/EXEE/XPSE/XCPR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U29579; AAA69271.1; -;  
 CC EMBL; A5000359; AAC75803.1; -;  
 CC EMBL; Y07525; -; NOT\_ANNOTATED\_CDS.  
 CC EcoGene; EG12634; YGCB.  
 CC InterPro; IPR001410; DEAD.  
 CC InterPro; IPR001482; GSPII\_E.  
 CC InterPro; IPR001650; Helicase\_C.  
 CC Pfam; PF00270; DEAD; 1.  
 CC SMART; SM00490; HELIC; 1.  
 CC PROSITE; PS00662; T2SP\_E; FALSE\_NEG.  
 KW Hypothetical protein; Transport; ATP-binding; Complete proteome.  
 FT NP\_BIND 314 321 ATP (POTENTIAL).  
 FT CONFLICT 118 118 G -> R (IN REF. 2).  
 FT CONFLICT 334 335 QQ -> PL (IN REF. 2).  
 SQ SEQUENCE 888 AA; 100544 MW; 4CA3F5371B1BF0F2 CRC64;

Query Match 41.8%; Score 38; DB 1; Length 888;  
 Best Local Similarity 38.5%; Pred. No. 78;  
 Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 5 YSIGRYSVRFWSK 17  
 I: | | | | | | | | | |  
 Db 74 HDIGKFDIRFQYK 86

RESULT 11

Wed Feb 13 07:52:32 2002

us-09-485-571-30.rsp

```

AXN1_HUMAN STANDARD; PRT; 900 AA.
ID AXN1_HUMAN
AC 015169;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AXIN1 (AXIS INHIBITION PROTEIN 1) (HAXIN) (FRAGMENT).
GN AXIN1 OR AXIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=9230313;
RX MEDLINE=97373830; Zhang T., Hsu W., Vasicek T.J., Perry W.L. III,
RA Zeng L., Fagotto F., Gumbiner B.M., Constantini F.;
RA Lee J.J., Tilghman S.M., Gumbiner B.M., Constantini F.;
RT "The mouse Fused locus encodes Axin, an inhibitor of the Wnt signaling
RT pathway that regulates embryonic axis formation.";
RL Cell 90:181-192(1997).
CC -!- FUNCTION: INHIBITOR OF THE WNT SIGNALING PATHWAY. DOWN-REGULATES
CC BETA-CATENIN. PROBABLY FACILITATE THE PHOSPHORYLATION OF BETA-
CC CATEIN AND APC BY GSK-3B.
CC -!- SUBUNIT: INTERACTS WITH GLYCOGEN SYNTHASE KINASE-3 BETA (GSK-3B)
CC AND BETA-CATENIN. THE INTERACTION BETWEEN AXIN AND BETA-CATENIN
CC OCCURS VIA THE ARMADILLO REPEATS CONTAINED IN BETA-CATENIN.
CC TERNARY COMPLEX. MAY ALSO BINDS TO PLAKOGLOBIN (GAMMA-CATENIN),
CC APC, DVL AND PP2A.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC -!- PTM: PROBABLY PHOSPHORYLATED BY GSK-3B AND DEPHOSPHORYLATED BY
CC PP2A.
CC -!- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DIX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF009674; AAC51624.1;
CC HSP: P49799; IAGR.
CC MIM: 603816;
CC InterPro: IPR001158; DIX.
CC Pfam: PF00778; DIX; 1.
CC Pfam: PF00615; RGS; 1.
CC ProDom: PD001580; RGS; 1.
CC ProDom: PD003639; DIX; 1.
CC SMART: SM00021; DAX; 1.
CC SMART: SM00315; RGS; 1.
CC PROSITE: PS50132; RGS; 1.
CC Developmental protein; Phosphorylation.
FT NON_TER 1
FT DOMAIN 125 248 RGS.
FT DOMAIN 385 470 GSK-3B BINDING SITE (BY SIMILARITY).
FT DOMAIN 471 540 BETA-CATENIN BINDING SITE (BY
FT DOMAIN 818 900 SIMILARITY).
FT DOMAIN 818 900 DIX.
SQ SEQUENCE 900 AA; 99803 MW; E5F990B11FC7B3B CRC64;

Query Match 41.8%; Score 38; DB 1; Length 900;
Best Local Similarity 56.2%; Pred. No. 79;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 SRRYSIGRYSVRFWSK 17
DB 320 SRRYSEGREFRYGSWR 335

RESULT 12
BCHH_RHOSH STANDARD; PRT; 1193 AA.
ID BCHH_RHOSH
AC Q9RFD5;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAGNESIUM-CHELATASE SUBUNIT H (MG-PROTOPORPHYRIN IX CHELATASE SUBUNIT
DE H).
GN BCHH.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=20115911; PubMed=10648776;
RA Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT sphaeroides 2.4.1.";
RL Nucleic Acids Res. 28:862-867(2000).
CC -!- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES
CC A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-
CC PROTOPORPHYRIN IX.
CC -!- PATHWAY: BACTERIOCHLOROPHYLL BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNIT H FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF195122; AAF24273.1;
CC InterPro: IPR003672; CbnN/Mg_chelatase.
CC Pfam: PF02514; cbnN-Mg_chel; 1.
CC Photosynthesis; Chlorophyll biosynthesis.
KW SEQUENCE 1193 AA; 129205 MW; 13DDEBD375223151 CRC64;

Query Match 41.8%; Score 38; DB 1; Length 1193;
Best Local Similarity 46.2%; Pred. No. 11e+02;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 RYSIGRYSVRFWSW 16
DB 193 RYLSVRSYSANRAW 205

RESULT 13
ATC8_YEAST STANDARD; PRT; 1656 AA.
ID ATC8_YEAST
AC Q12674;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE 5 (EC 3.6.3.13).
GN YMR162C OR YMB520.11C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
CC OF ATP COUPLED WITH THE TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
CC

```

CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
(E1-E2 ATPASES). SUBFAMILY IV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z49705; CAA89798.1; -;  
DR SGD; S0004772; YMR162C.  
DR InterPro; IPR001757; E1-E2 ATPase.  
DR InterPro; IPR001454; Hydrolase.  
DR Pfam; PF00702; Hydrolase; 1.  
DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
KW Hypothetical protein; Hydrolase; Transmembrane; Phosphorylation;  
KW Magnesium; ATP-binding.  
FT TRANSMEM 165 185 POTENTIAL.  
FT TRANSMEM 452 472 POTENTIAL.  
FT TRANSMEM 496 516 POTENTIAL.  
FT TRANSMEM 1158 1178 POTENTIAL.  
FT TRANSMEM 1319 1339 POTENTIAL.  
FT TRANSMEM 1366 1386 POTENTIAL.  
FT TRANSMEM 1396 1416 POTENTIAL.  
FT TRANSMEM 1433 1453 POTENTIAL.  
FT TRANSMEM 1474 1494 POTENTIAL.  
FT MOD\_RES 566 566 PHOSPHORYLATION (PROBABLE).  
SQ SEQUENCE 1656 AA; 188318 MW; A20A823BEB401184 CRC64;  
  
Query Match 41.88; Score 38; DB 1; Length 1656;  
Best Local Similarity 58.3%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 5 YSIGRYSVRFWSW 16  
I I I I I I I  
Db 1355 YSIGRLSQGFNN 1366  
  
RESULT 14  
VMSA\_HPBVT STANDARD; PRT; 400 AA.  
AC Q05496;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE MAJOR SURFACE ANTIGEN PRECURSOR.  
GN S.  
OS Hepatitis B virus (subtype adw4 / strain Brazil / isolate w4B).  
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=45410;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93346970; PubMed=8345355;  
RA Naumann H., Schaefer S., Yoshida C.F.T., Gaspar A.M.C., Repp R.,  
RA Gerlich W.H.;  
RT "Identification of a new hepatitis B virus (HBV) genotype from Brazil  
RT that expresses HBV surface antigen subtype adw4.";  
RL J. Gen. Virol. 74:1627-1632(1993).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X69798; CAA49455.1; -;

DR InterPro; IPR000349; Hepadnavir\_surfag.  
DR Pfam; PF00695; VMSA; 1.  
KW Antigen.  
FT PROPEP 1 174  
FT CHAIN 175 400 MAJOR SURFACE ANTIGEN.  
SQ SEQUENCE 400 AA; 43551 MW; 4A5A2212E4B3E117 CRC64;  
  
Query Match 41.2%; Score 37.5; DB 1; Length 400;  
Best Local Similarity 41.2%; Pred. No. 41;  
Matches 7; Conservative 4; Mismatches 1; Indels 5; Gaps 1;  
  
Qy 5 YSIGRY-----SVRFWSW 16  
:::|:| | | | |  
Db 330 WALGKYLWEMASARFSW 346  
  
RESULT 15  
VMSA\_PYRFU STANDARD; PRT; 648 AA.  
AC P49057;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ALPHA-AMYLASE (EC 3.2.1.1).  
GN VMSA.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=94043280; PubMed=8226990;  
RA Laderman K.A., Asada K., Uemori T., Mukai H., Taguchi Y., Kato I.,  
RA Anfinsen C.B.;  
RT "Alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus  
RT furiosus. Cloning and sequencing of the gene and expression in  
RT Escherichia coli.";  
RL J. Biol. Chem. 268:24402-24407(1993).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=94043279; PubMed=8226989;  
RA Laderman K.A., Davis B.R., Krutzsch H.C., Lewis M.S., Griko Y.V.,  
RA Privalov P.L., Anfinsen C.B.;  
RT "The purification and characterization of an extremely thermostable  
RT alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus  
RT furiosus";  
RL J. Biol. Chem. 268:24394-24401(1993).  
CC -!- FUNCTION: DISPLAYS A BROAD RANGE OF SUBSTRATE SPECIFICITY, WITH  
CC THE CAPACITY TO HYDROLYZE CARBOHYDRATES AS SIMPLE AS MALTOPIROSE.  
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC  
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.  
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- MISCELLANEOUS: THE ISOELECTRIC POINT IS 4.3. THE ENZYME DISPLAYS  
CC OPTIMAL ACTIVITY, WITH SUBSTANTIAL THERMAL STABILITY, AT 100  
CC DEGREES CELSIUS, WITH THE ONSET OF ACTIVITY AT APPROXIMATELY 40  
CC DEGREES CELSIUS.  
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; I22346; AAA72035.1; -;  
KW Hydrolase; Glycosidase; Carbohydrate metabolism.  
FT INT\_MT 0  
SQ SEQUENCE 648 AA; 76178 MW; BF7A495F084E0FB1 CRC64;

us-09-485-571-30.rsp

Wed Feb 13 07:52:32 2002

Query Match 41.2%; Score 37.5; DB 1; Length 648;  
Best Local Similarity 26.3%; Pred. No. 68;  
Matches 5; Conservative 8; Mismatches 1; Indels 5; Gaps 1;

QY 4 RYSGRYSVRES-----WK 17  
:|:|:|:|:|:|  
DB 587 KYAVGKFALKEDEMEVWK 605

Search completed: February 12, 2002, 12:39:54  
Job time: 807 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:42 ; Search time 232.64 Seconds  
(without alignments)  
10.689 Million cell updates/sec

Title: US-09-485-571-30

Perfect score: 91

Sequence: 1 RRRYSIGRYSVRESWK 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_invertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	49.5	170	12 Q66944	Q66944 feline immu
2	45	49.5	208	12 Q89177	Q89177 feline immu
3	44	48.4	124	10 Q9SNA5	Q9SNA5 arabidopsis
4	43	47.3	170	12 Q66942	Q66942 feline immu
5	43	47.3	675	2 Q9AJ16	Q9AJ16 escherichia
6	43	47.3	855	12 Q66961	Q66961 feline immu
7	42	46.2	540	10 Q9LY83	Q9LY83 arabidopsis
8	42	46.2	855	12 Q66962	Q66962 feline immu
9	41	45.1	225	12 Q85539	Q85539 feline immu
10	41	45.1	717	4 Q9YA42	Q9YA42 homo sapien
11	41	45.1	734	4 Q9H9V5	Q9H9V5 homo sapien
12	41	45.1	1193	10 Q9MA43	Q9MA43 arabidopsis
13	41	45.1	1235	12 Q9IBP3	Q9IBP3 western equ
14	41	45.1	1236	12 Q9J1K1	Q9J1K1 western equ
15	41	45.1	1265	3 Q9UVD8	Q9UVD8 alternaria
16	40.5	44.5	625	5 Q9GUT8	Q9GUT8 leishmania
17	40	44.0	208	12 Q89185	Q89185 feline immu
18	40	44.0	208	12 Q89186	Q89186 feline immu
19	40	44.0	223	12 Q86106	Q86106 feline immu

20 40 44.0 224 12 Q85536  
21 40 44.0 240 2 Q45282  
22 40 44.0 242 12 P87681  
23 40 44.0 267 2 Q9ZKK2  
24 40 44.0 273 12 P89340  
25 40 44.0 331 5 Q9XUV8  
26 40 44.0 363 4 Q9PIB2  
27 40 44.0 401 2 Q9R881  
28 40 44.0 723 10 Q9M1H3  
29 40 44.0 816 10 Q9LJG4  
30 40 44.0 895 12 Q9LJX3  
31 39 42.9 63 12 Q9Q662  
32 39 42.9 117 1 Q9HJK9  
33 39 42.9 170 12 Q66940  
34 39 42.9 170 12 Q66947  
35 39 42.9 171 10 Q9ZUG0  
36 39 42.9 171 12 Q12123  
37 39 42.9 191 1 Q58921  
38 39 42.9 222 5 Q9NEF4  
39 39 42.9 226 12 Q67922  
40 39 42.9 261 10 Q9S9X1  
41 39 42.9 281 12 Q67927  
42 39 42.9 291 10 Q9FTM7  
43 39 42.9 303 5 Q9W4Y0  
44 39 42.9 353 5 P91233  
45 39 42.9 379 10 Q96498

#### ALIGNMENTS

RESULT 1  
Q66944 ID Q66941 PRELIMINARY; PRT; 170 AA.  
AC Q66944;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE GP100 SURFACE PROTEIN (FRAGMENT).  
GN ENV.  
OS Feline immunodeficiency virus.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11673;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NORRISTOWN, PA;  
RX MEDLINE=97286227; PubMed=9151811;  
RA Bachmann M.H., Mathiason-Dubard C., Learn G.H., Rodrigo A.G.,  
RA Sodora D.L., Mazzetti P., Hoover E.A., Mullins J.I.;  
RT "Genetic diversity of feline immunodeficiency virus: dual infection,  
RT recombination, and distinct evolutionary rates among envelope sequence  
RT clades.";  
RL J. Virol. 71:4241-4253(1997).  
DR EMBL; J57018; AAC57313.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 170 170  
SQ SEQUENCE 170 AA; 20078 MW; 7E68D346AF26CCCB CRC64;  
  
Query Match 49.5%; Score 45; DB 12; Length 170;  
Best Local Similarity 43.8%; Pred. No. 6.4;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 RRRYSIGRYSVRESWK 16  
:|:|:| |:|:| |  
Db 92 KKRHSEGRFRICRW 107  
  
RESULT 2  
Q89177 ID Q89177 PRELIMINARY; PRT; 208 AA.  
AC Q89177;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)



```

Db 563 RHILGRYSVR 573
      ||:|||||
      3 RRYISGRYSVRSWK 17
      |||||
      513 RDSIGEYSANDAW 527

RESULT 6
Q66961 ID Q66961 PRELIMINARY; PRT; 855 AA.
AC Q66961;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENVELOPE POLYPROTEIN.
GN ENV.
OS Feline immunodeficiency virus.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11673;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AOMORI 1;
RX MEDLINE=95264455; PubMed=7745712;
RA Kakinuma S., Motokawa K., Hohdatsu T., Yamamoto J., Koyama H.,
RA Hashimoto H.;
RT "Nucleotide sequence of feline immunodeficiency virus: classification
RT of Japanese isolates into two subtypes which are distinct from non-
RT Japanese subtypes.";
RL J. Virol. 69:3639-3646(1995).
DR EMBL; D37816; BAA07062.1; -
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Polyprotein; Envelope protein.
SQ SEQUENCE 855 AA; 97876 MW; 1FDD34E0DACFE204 CRC64;

Query Match 47.3%; Score 43; DB 12; Length 855;
Best Local Similarity 43.8%; Pred. No. 80;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSRYSIGRYSVRSWK 16
      |||||
      449 KSRHSEARFRIRCKW 464

Db 563 RHILGRYSVR 573

RESULT 7
Q9LY83 ID Q9LY83 PRELIMINARY; PRT; 540 AA.
AC Q9LY83;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TRNA SYNTHASE-LIKE PROTEIN.
GN F18022.250.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL163817; CAB87785.1; -
DR InterPro; IPR002501; TruB_N.
DR Pfam; PF01509; TruB_N; 1.
SQ SEQUENCE 540 AA; 61475 MW; A9379FFDB44ED803 CRC64;

Query Match 46.2%; Score 42; DB 10; Length 540;
Best Local Similarity 53.3%; Pred. No. 71;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 RRYISGRYSVRSWK 17
      |||||
      513 RDSIGEYSANDAW 527

RESULT 8
Q66962 ID Q66962 PRELIMINARY; PRT; 855 AA.
AC Q66962;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENVELOPE POLYPROTEIN.
GN ENV.
OS Feline immunodeficiency virus.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11673;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AOMORI 2;
RX MEDLINE=95264455; PubMed=7745712;
RA Kakinuma S., Motokawa K., Hohdatsu T., Yamamoto J., Koyama H.,
RA Hashimoto H.;
RT "Nucleotide sequence of feline immunodeficiency virus: classification
RT of Japanese isolates into two subtypes which are distinct from non-
RT Japanese subtypes.";
RL J. Virol. 69:3639-3646(1995).
DR EMBL; D37817; BAA07063.1; -
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Polyprotein; Envelope protein.
SQ SEQUENCE 855 AA; 98196 MW; 1158C3FA8DA06BED CRC64;

Query Match 46.2%; Score 42; DB 12; Length 855;
Best Local Similarity 37.5%; Pred. No. 1.2e+02;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSRYSIGRYSVRSWK 16
      |||||
      449 KSKYSEARFRIRCKW 464

Db 563 RHILGRYSVR 573

RESULT 9
Q85539 ID Q85539 PRELIMINARY; PRT; 225 AA.
AC Q85539;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE GPI00 (FRAGMENT).
GN ENV.
OS Feline immunodeficiency virus.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11673;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USOKLGRLO2B;
RX MEDLINE=94187063; PubMed=8139008;
RA Sodora D.L., Shpaer E.G., Kitchell B.E., Dow S.W., Hoover E.A.,
RA Mullins J.I.;
RT "Identification of three feline immunodeficiency virus (FIV) env gene
RT subtypes and comparison of the FIV and human immunodeficiency virus
RT type 1 evolutionary patterns.";
RL J. Virol. 68:2230-2238(1994).
DR EMBL; U02421; AAA18045.1; -
FT NON_TEF 1
FT NON_TEF 225
SQ SEQUENCE 225 AA; 26039 MW; C33E18FFB5458C58 CRC64;

```

Wed Feb 13 07:52:32 2002

us-09-485-571-30.ispt

Query Match 45.1%; Score 41; DB 12; Length 225;  
Best Local Similarity 37.5%; Pred. No. 41;  
Matches 6; Conservative 5; Mismatches 0; Gaps 0;

OY 1 SRRYSIGRYSVRFWSW 16  
Db 107 KSKQSEARFRICKW 122

RESULT 10  
OY4W2 PRELIMINARY; PRT; 717 AA.  
AC OY4W2;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE DJ475B7.2 (NOVEL PROTEIN).  
GN DJ475B7.2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Howden P.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL050306; CAB51351.1; -. 651B7F0FFB3FD07F CRC64;  
SQ SEQUENCE 717 AA; 81242 MW; 651B7F0FFB3FD07F CRC64;

Query Match 45.1%; Score 41; DB 4; Length 717;  
Best Local Similarity 37.5%; Pred. No. 1.4e+02;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 2 SRRYSIGRYSVRFWSW 17  
Db 420 ARFSAGQWEARRGWR 435

RESULT 11  
OY9HV5 PRELIMINARY; PRT; 734 AA.  
AC OY9HV5;  
DT 01-NAR-2001 (TREMBlrel. 16, Created)  
DT 01-NAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-NAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CDNA FLJ12525 F1S, CLONE NT2RM400030, WEAKLY SIMILAR TO LAS1 PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwananagi T.;  
RP "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK022587; BAB14114.1; -. A35CC38F95C39F7D CRC64;  
SQ SEQUENCE 734 AA; 83064 MW; . A35CC38F95C39F7D CRC64;

Query Match 45.1%; Score 41; DB 4; Length 734;  
Best Local Similarity 37.5%; Pred. No. 1.5e+02;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 2 SRRYSIGRYSVRFWSW 17  
Db 437 ARFSAGQWEARRGWR 452

RESULT 12  
OYMA43 PRELIMINARY; PRT; 1193 AA.  
ID OYMA43;  
AC OYMA43;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE T2OM3.10 PROTEIN.  
GN T2OM3.10.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Liu S.X., Sakano H., Yu G., Toriumi M., Lenz C., Lee J.M., Li J.,  
RA Liu A., Gonzalez A., Liu K., Vaysberg M., Chin C., Wong B., Choi E.,  
RA Chiou J., Pham P., Koo T., Altafi H., Brooks S., Buehler E., Chao Q.,  
RA Conn L., Conway A.B., Hansen N., Johnson-Hopson C., Khan S., Kim C.,  
RA Lam B., Nguyen M., Palm C., Shinn P., Tambunga G., Davis R.W.,  
RA Becker J.R., Federspiel N.A., Theologis A.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC009999; AAF29390.1; -.  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR InterPro; IPR003889; FYrich\_C.  
DR InterPro; IPR003888; FYrich\_N.  
DR InterPro; IPR001965; PHD.  
DR InterPro; IPR003616; PostSET.  
DR InterPro; IPR000313; PWWP.  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00855; PWWP; 1.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00542; FYRC; 1.  
DR SMART; SM00541; FYRN; 1.  
DR SMART; SM00249; PHD; 2.  
DR SMART; SM00508; PostSET; 1.  
DR SMART; SM00293; PWWP; 1.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
DR PROSITE; PS0280; SET; 1.  
SQ SEQUENCE 1193 AA; 135416 MW; 4F2B3D5776A30DF6 CRC64;

Query Match 45.1%; Score 41; DB 10; Length 1193;  
Best Local Similarity 39.1%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 4; Indels 8; Gaps 1;

OY 3 RRYSIGRYS-----VRFWSW 17  
Db 564 RYSSGKYQDHPTGYRVRVWVK 586

RESULT 13  
OYIBP3 PRELIMINARY; PRT; 1235 AA.  
ID OYIBP3;  
AC OYIBP3;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE STRUCTURAL POLYPROTEIN.  
OS Western equine encephalitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=11039;  
RN [1]  
RP SEQUENCE OF 1-259 FROM N.A.  
RC STRAIN=MCILLAN;  
RX MEDLINE=96097271; PubMed=8535272;

RA Uryvaev L.V., Lebedev A.Iu, Sokolova T.M., Iuferov V.P.;  
 RT "Primary structure of the nucleocapsid gene C and the protein coded by  
 RL it from the Western equine encephalomyelitis virus.";  
 RN Dokl. Akad. Nauk 344:397-401(1995).  
 RP (2)  
 RC SEQUENCE OF 1-259 FROM N.A.  
 RC STRAIN-MCMILLAN;  
 RX MEDLINE-97130212; PubMed-8999681;  
 RA Uryvaev L.V., Lebedev A.Iu.;  
 RT "Comparative analysis of primary structure of nucleocapsid protein  
 RL from Western equine encephalomyelitis virus and other alphaviruses.";  
 RN Vopr. Virusol. 41:252-259(1996).  
 RP (3)  
 RC SEQUENCE OF 320-742 FROM N.A.  
 RC STRAIN-MCMILLAN;  
 RX MEDLINE-98100973; PubMed-9471275;  
 RA Uryvaev L.V., Lebedev A.Iu, Sokolova T.M., Iuferov V.P.;  
 RT "Capsid glycoprotein of E2 Encephalitis Virus, Western Equine: primary  
 RL structure of gene and its product.";  
 RN Dokl. Akad. Nauk 357:134-139(1997).  
 RP (4)  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-MCMILLAN;  
 RA Uryvaev L.V.;  
 RT "Primary structure of WEEV 26S RNA.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF229608; AAF60166.1; -;  
 DR InterPro; IPR001836; Alpha\_core.  
 DR InterPro; IPR002548; Alpha\_E1\_glycop.  
 DR InterPro; IPR000936; Alpha\_E2\_glycop.  
 DR InterPro; IPR002533; Alpha\_E3\_glycop.  
 DR InterPro; IPR000930; Togavirin.  
 DR Pfam; PF00944; Alpha\_core; 1.  
 DR Pfam; PF01589; Alpha\_E1\_glycop; 1.  
 DR Pfam; PF00943; Alpha\_E2\_glycop; 1.  
 DR Pfam; PF01563; Alpha\_E3\_glycop; 1.  
 DR PRINTS; PR00798; TOGAVIRIN.  
 KW Polyprotein.  
 FT CHAIN 1 259 C-PROTEIN.  
 FT CHAIN 260 319 E3-POLYPEPTIDE.  
 FT CHAIN 320 742 E2-PROTEIN.  
 FT CHAIN 743 807 6K-POLYPEPTIDE.  
 FT CHAIN 808 1235 E1-PROTEIN.  
 SQ SEQUENCE 1235 AA; 136352 MW; 27AD7CED71EEBD4E CRC64;

Query Match 45.1%; Score 41; DB 12; Length 1235;  
 Best Local Similarity 40.0%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SRRYSIGRYSVRFWSW 16  
 ||:|:|:|:|:|  
 Db 635 SRNFSVGRGLEYYVW 649

RESULT 14  
 Q9J1K1 PRELIMINARY; PRT; 1236 AA.  
 AC Q9J1K1;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE STRUCTURAL POLYPEPTIDE.  
 OS Western equine encephalitis virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 CC Alphavirus.  
 OX NCBI\_TaxID=11039;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-71V-1658;  
 RX MEDLINE-20109043; PubMed-10640553;  
 RA Necolitzky D.J., Schmalz F.L., Parker M.D., Rayner G.A., Fisher G.R.,  
 RA Trent D.W., Bader D.E., Nagata L.P.;

RT "Complete genomic RNA sequence of western equine encephalitis virus  
 RL and expression of the structural genes.";  
 RN J. Gen. Virol. 81:151-159(2000).  
 DR EMBL; AF214040; AAF28340.1; -;  
 DR InterPro; IPR001836; Alpha\_core.  
 DR InterPro; IPR002548; Alpha\_E1\_glycop.  
 DR InterPro; IPR000936; Alpha\_E2\_glycop.  
 DR InterPro; IPR002533; Alpha\_E3\_glycop.  
 DR InterPro; IPR000930; Togavirin.  
 DR Pfam; PF00944; Alpha\_core; 1.  
 DR Pfam; PF01589; Alpha\_E1\_glycop; 1.  
 DR Pfam; PF00943; Alpha\_E2\_glycop; 1.  
 DR Pfam; PF01563; Alpha\_E3\_glycop; 1.  
 DR PRINTS; PR00798; TOGAVIRIN.  
 KW Polyprotein.  
 SQ SEQUENCE 1236 AA; 136266 MW; 56B752C0D19CD3F5 CRC64;

Query Match 45.1%; Score 41; DB 12; Length 1236;  
 Best Local Similarity 40.0%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SRRYSIGRYSVRFWSW 16  
 ||:|:|:|:|:|  
 Db 635 SRNFSVGRGLEYYVW 649

RESULT 15  
 Q9UVD8 PRELIMINARY; PRT; 1265 AA.  
 AC Q9UVD8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE POL (FRAGMENT).  
 GN POL.  
 OS Alternaria alternata.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;  
 OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.  
 OX NCBI\_TaxID=5599;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-15A; TRANSPOSON-LTR-RETROTRANSPOSON REAL;  
 RA Tsuge T., Kaneko I.;  
 RT "REAL: an LTR-retrotransposon of the plant pathogenic fungus  
 RL Alternaria alternata.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SEMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE  
 CC TRANSCRIPTASE).  
 DR EMBL; AB025309; BAA89272.1; -;  
 DR InterPro; IPR000953; Chromo.  
 DR InterPro; IPR000504; REM.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR InterPro; IPR000737; Squash.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00078; rvt; 1.  
 DR PRINTS; PR00293; SQUASHINHBTR.  
 DR SMART; SM00298; CHROMO; 1.  
 DR PROSITE; PS00013; CHROMO\_2; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 KW RNA-directed DNA polymerase.  
 FT NON\_TER 1  
 SQ SEQUENCE 1265 AA; 144367 MW; 6E5DAF287D461C6D CRC64;

Query Match 45.1%; Score 41; DB 3; Length 1265;  
 Best Local Similarity 46.7%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 RRYSIGRYSVRFWSW 17  
 :|:|:|:|:|:|  
 Db 1116 QOYSTGRSSKKLDWK 1130

us-09-485-571-30.rspt

Wed Feb 13 07:52:32 2002

Search completed: February 12, 2002, 12:38:43  
Job time: 756 sec

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:30 ; Search time 242.57 Seconds  
(without alignments)  
5.497 Million cell updates/sec

Title: US-09-485-571-20

Perfect score: 33

Sequence: 1 XGGXXXXXXXXXXXXXG 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	51.5	738	19	AAW56163 New DNA sequence i
2	17	51.5	803	21	AAW00066 KIAA0160 polypepti
3	16	48.5	102	21	AAW02285 Human secreted pro
4	16	48.5	129	21	AAW32776 Eucalyptus grandis
5	16	48.5	394	21	AAW79674 Human potassium ch
6	16	48.5	334	21	AAW87291 Human signal pepti
7	16	48.5	526	21	AAW84906 A human proliferat
8	16	48.5	526	22	AAW94371 Human protein sequ
9	16	48.5	623	20	AAW02540 Protein encoded by
10	16	48.5	736	12	AAW13949 SUP-B27 t(1;19) tr
11	16	48.5	742	12	AAW15158 E2A/pr1 fusion pro

12	16	48.5	819	12	AAW13948 SUP-B27 t(1;19) tr
13	16	48.5	825	12	AAW13951 E2A/pr1 fusion pro
14	15	45.5	104	21	AAW02803 Human secreted pro
15	15	45.5	256	20	AAW02538 Protein encoded by
16	15	45.5	231	22	AAW82608 Spider recombinant
17	15	45.5	349	20	AAW09298 Rice beta-glucanas
18	15	45.5	377	20	AAW09306 Rice Gln9 CDS prot
19	15	45.5	503	22	AAW09306 C glutamicum prote
20	15	45.5	630	20	AAW02541 Protein encoded by
21	15	45.5	681	22	AAW82609 Spider recombinant
22	15	45.5	691	22	AAW82610 Spider recombinant
23	14	42.4	90	21	AAW53989 Human colon cancer
24	14	42.4	111	19	AAW61478 A. fumigatus aller
25	14	42.4	126	22	AAW93287 Human protein HP10
26	14	42.4	142	21	AAW93287 Arabidopsis thalia
27	14	42.4	146	21	AAW93287 Arabidopsis thalia
28	14	42.4	152	21	AAW93287 Arabidopsis thalia
29	14	42.4	161	21	AAW93287 Arabidopsis thalia
30	14	42.4	180	22	AAW40773 Human polypeptide
31	14	42.4	180	22	AAW40773 Human polypeptide
32	14	42.4	362	20	AAW06324 Mouse G protein co
33	14	42.4	362	20	AAW06324 Rat MACHR-6 protei
34	14	42.4	370	20	AAW92977 Rat MACHR-6 protei
35	14	42.4	409	22	AAW93991 Human protein sequ
36	14	42.4	422	21	AAW41764 Arabidopsis thalia
37	14	42.4	445	20	AAW06323 Rat G protein coup
38	14	42.4	445	21	AAW15382 Rat G-protein coup
39	14	42.4	451	19	AAW71373 Human HKA42 poly
40	14	42.4	451	19	AAW71373 Partial human HKA4
41	14	42.4	501	22	AAW38988 Human polypeptide
42	14	42.4	568	21	AAW41763 Arabidopsis thalia
43	14	42.4	579	22	AAW93285 Human protein HP10
44	14	42.4	590	21	AAW41762 Arabidopsis thalia
45	14	42.4	605	21	AAW50293 Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAW56163  
ID AAW56163 standard; Protein; 738 AA.  
XX  
AC AAW56163;  
XX  
XX  
DT 28-JUL-1998 (first entry)  
XX  
DE New DNA sequence isolated from Pinctada fucata.  
XX  
KW Pinctada fucata; protein; cosmetic.  
XX  
OS Pinctada fucata.  
XX

PN JPL0080285-A.  
XX  
XX  
XX 31-MAR-1998.  
XX  
XX 28-MAY-1997; 97JP-0138461.  
XX  
XX 15-JUL-1996; 96JP-0184459.  
XX  
XX (MIKI-; MIKIMOTO SEIYAKU KK.  
XX  
XX WPI; 1998-254410/23.  
XX  
XX N-PSDB; AAW22683.  
XX  
PT New cDNA and e.g. vector, host cell and polypeptide - used to  
XX produce polypeptide in high yields, which is used in cosmetics  
XX  
XX Claim 9; Pages 9-11; 15pp; Japanese.  
XX  
XX The present sequence represents protein encoded by a new DNA sequence  
XX isolated from Pinctada fucata. The protein be used as an ingredient  
CC

```
CC in cosmetics.
XX Sequence 738 AA;
SQ

Query Match 51.5%; Score 17; DB 19; Length 738;
Best Local Similarity 17.6%; Pred. No. 1e+04;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXXXXXXXG 18
  ||
Db 280 ggsaaaaaaaagg 296

RESULT 2
AAB00066
ID AAB00066 standard; Protein: 803 AA.
XX
AC AAB00066;
XX
DT 16-NOV-2000 (first entry)
XX
DE KIAA0160 polypeptide.
XX
KW Vernalization gene; VRN2; plant characteristic; flowering time;
KW leaf size; leaf shape; shade avoidance response; reproduction;
KW breeding; pollination; cultivation; human.
XX
OS Homo sapiens.
XX
PN WO200044918-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-GB00248.
XX
PR 28-JAN-1999; 99GB-0001927.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI Dean C, Gendall A;
XX
WPI: 2000-499333/44.
DR N-PSDB; AAA47759.
XX
PT Isolated vernalization gene VRN2 is used to produce transgenic plants
PT with altered vernalization response, flowering time, leaf size and/or
PT shape or shade avoidance response for maximized reproductive success
XX
PS Disclosure; Page 79; 105pp; English.
XX
CC Isolated nucleic acid sequences obtained from the VRN2 locus of a
CC plant encode polypeptides which are capable of affecting one or more
CC vernalization responses such as, flowering time, leaf size and/or
CC shape or the shade avoidance response of a plant into which the
CC nucleic acid is introduced. Introducing such sequences into plants
CC to alter these characteristics maximises the reproductive success of
CC the plant. This protein is encoded an isolated human sequence which
CC has homology to the Arabidopsis thaliana VRN2 gene over a short region
XX
XX near the N-terminus.
XX
SQ Sequence 803 AA;

Query Match 51.5%; Score 17; DB 21; Length 803;
Best Local Similarity 17.6%; Pred. No. 1.1e+04;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXXXXXXXG 18
  ||
Db 113 ggsysassssaaaaag 129

us-09-485-571-20.rag

RESULT 3
AAG02285
ID AAG02285 standard; Protein: 102 AA.
XX
AC AAG02285;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 6366.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
WPI: 2000-500381/45.
DR N-PSDB; AAC02291.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 6366; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX were prepared from total human RNAs or polyA+ RNAs derived from 30
XX different tissues. EST sequences usually correspond mainly to the 3'
XX untranslated region (UTR) of the mRNA because they are often obtained
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX those cases where longer cDNA sequences have been obtained, the full 5'
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX ends and can therefore be used to obtain full length cDNAs and genomic
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX chromosome mapping procedures. They are used to obtain upstream
XX regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 102 AA;

Query Match 48.5%; Score 16; DB 21; Length 102;
Best Local Similarity 17.6%; Pred. No. 4.5e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXXXXXXXG 18
  ||
Db 23 ggsstgtttttttgg 39

RESULT 4
AAB32776
ID AAB32776 standard; Protein: 129 AA.
XX
AC AAB32776;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor protein sequence #234.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
```



KW poplar; sweetgum; teak; mahogany; bzIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;  
 XX type 2 Cys2His2; CCAAT box element; MYB.  
 OS Eucalyptus grandis.  
 XX  
 PN WO200053724-A2.  
 XX  
 XX 14-SEP-2000.  
 XX  
 XX 09-MAR-2000; 2000WO-US06112.  
 XX  
 PR 11-MAR-1999; 99US-0266513.  
 PR 18-AUG-1999; 99US-0149485.  
 XX  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 XX Wood M, McGrath A, Shenk MA, Glenn M;  
 XX WPI; 2000-579369/54.  
 DR  
 XX  
 XX New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT .  
 XX  
 XX Claim 8; Page 300; 747pp; English.  
 PS  
 XX The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
 CC transcription factor. The transcription factor may be used to produce a  
 CC plant having modified gene expression such as a woody plant e.g. a  
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
 CC to modify the activity of a polypeptide in a plant. The transcription  
 CC factors of the present invention are members from the following families  
 CC of regulatory proteins: bzIP, bzIP family of G-box binding factors, basic  
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
 CC zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2  
 CC Cys2His2, CCAAT box elements and MYB.  
 XX  
 XX Sequence 129 AA;  
 SQ  
 Query Match 48.5%; Score 16; DB 21; Length 129;  
 Best Local Similarity 17.6%; Pred. No. 5.3e+03;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 GGXXXXXXXG 18  
 ||  
 Db 94 ggsssaarsasg 110  
 ||  
 RESULT 5  
 AAY79674  
 ID AAY79674 standard; Protein; 394 AA.  
 AC AAY79674;  
 XX  
 XX 29-AUG-2000 (first entry)  
 DT Human potassium channel TASK1.  
 DE  
 XX TASK1; TWIK-related acid-sensitive K+ channel 1; human;  
 KW potassium channel; drug screening; hypertension;  
 KW hypotensive; epilepsy; arrhythmia; vascular diseases;  
 KW neurodegenerative disease; ischaemia; anoxia; endocrine disease;  
 KW muscle disease; therapy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH

FT Modified-site 53 /note= "N-glycosylated"  
 FT Modified-site 323 /note= "O-phosphorylated"  
 FT Modified-site 383 /note= "O-phosphorylated"  
 FT Modified-site 392 /note= "O-phosphorylated"  
 FT Modified-site 393 /note= "O-phosphorylated"  
 FT Modified-site 393 /note= "O-phosphorylated"  
 XX  
 PN WO200027871-A2.  
 XX  
 XX 18-MAY-2000.  
 PD  
 XX 09-NOV-1999; 99WO-IB01886.  
 PF  
 XX 09-NOV-1998; 98US-0107692.  
 PR 08-NOV-1999; 99US-0436265.  
 PR  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA  
 XX Duprat F, Lesage F, Lazdunski M;  
 PI WPI; 2000-376487/32.  
 XX N-PSDB; AAA27746.  
 DR  
 XX New nucleic acid encoding a non-inactivating outwardly rectifying  
 PT potassium transport channel, designated TASK2, useful in the treatment  
 PT of hypertension or dysfunctions of the kidney, liver or pancreas -  
 PT Disclosure; Fig 8; 91pp; English.  
 PS  
 XX The present sequence is that of human TASK1 (TWIK-related  
 CC acid-sensitive K+ channel), a member of a new family of 2p  
 CC domain potassium channels, also including TWIK-1 (see AAY79673) and  
 CC novel TASK2 (see AAY79675). TASK1 is expressed in many different  
 CC tissues, and at particularly high levels in pancreas and placenta.  
 CC Host cells expressing TWIK-1 family members can be used to screen  
 CC for substances that modulate the activity of members of the TWIK-1  
 CC family of potassium channels. The drugs identified may be  
 CC useful in the treatment of diseases of the heart or of the nervous  
 CC system, such as epilepsy, arrhythmia, vascular diseases,  
 CC neurodegenerative diseases, kidney, liver or pancreas diseases,  
 CC hypertension, diseases associated with ischaemia or anoxia,  
 CC endocrine diseases associated with anomalies of hormone secretion,  
 CC and muscle diseases.  
 CC  
 SQ Sequence 394 AA;  
 Query Match 48.5%; Score 16; DB 21; Length 394;  
 Best Local Similarity 17.6%; Pred. No. 1.1e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 GGXXXXXXXG 18  
 ||  
 Db 276 ggsahttdtasaaag 292  
 ||  
 RESULT 6  
 AAY87291  
 ID AAY87291 standard; Protein; 394 AA.  
 XX  
 XX AAY87291;  
 AC  
 XX 11-MAY-2000 (first entry)  
 DT Human signal peptide containing protein HSPP-68 SEQ ID NO:68.  
 XX  
 DE Human; signal peptide-containing protein; HSPP; diagnosis; cancer;  
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
 KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;  
 KW

antiasthmatic; gene therapy; cell proliferation; neurological disorder;  
 reproductive disorder; developmental disorder; arteriosclerosis;  
 cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
 asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
 parkinson's disease; Huntington's diseases; ovulatory defect;  
 muscular dystrophy.

xx Homo sapiens.

XX  
DN  
W020000610-A2.

XX  
DD  
06-TAN-2000XX  
CE - TUN - 1000.  
99WQ-11514484XX  
1000  
0800-0090762

PR 31-JUL-1998; 98US-0094983.

PR 11-DEC-1998; 98US-0112129.

XX  
P3 (TNCV - ) TNCYTE PHARM INC.

XXXX

PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy K, Hillman JD  
PI Bandman O;

WPT: 2000-160673/14;

DR N-PSDB; AA298176.

PT New human signal peptide-containing proteins useful in treatment  
PT prevention and diagnosis of e.g. cancer, inflammation and  
PT cardiovascular disease \*

XX  
PS  
claim 1. Page 207-208; 327pp; English.

AA298109 to AA298242 encode AY87224 to AY87357 which represent the human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic, neuroprotective, cardiovascular and antiasthmatic activities, and can be used in gene therapy. HSPs can be used to treat or prevent disorders associated with decreased activity or function of HSP. Antagonists of HSP are used to treat or prevent disorders associated with increased activity or function of HSP. Such diseases include cell proliferation (including cancer), inflammation, cardiovascular, neurological, reproductive or developmental disorders, (e.g. arteriosclerosis, cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia, asthma, Crohn's disease, Alzheimer's, Parkinson's or Huntington's ischaemic heart disease, ovulatory defects, muscular dystrophy). HSP nucleic acids can be used for the recombinant production of HSP, for detecting HSP in standard hybridisation and amplification assays (for diagnosis and monitoring), in gene therapy, as antisense, triplex-forming or ribozyme therapeutics, for detecting related sequences or genetic variations, and for chromosomal mapping. HSP are also used to raise specific antibodies (Ab) and to screen for agonists and antagonists (potential therapeutic agents). Ab are used to diagnose, or monitor, HSP-related diseases (in usual immunoassays), as therapeutic antagonists, in competitive drug screens, and for purification of HSP from natural sources.

XX  
Sequence 394 AA:

Query Match	48.5%	Score 16;	DB 21;	Length 394;
Best Local Similarity	17.6%	Pred.:No.	1.1e+04;	
Matches 3;	Conservative	0;	Mismatches 14;	Indels 0;
Gaps	0;			

QY 2 GGXXXXXXXXXXXXXG 18

pb 276 qqsahhttdtasstaaq 292

## RESULT 7

PR 22-APR-1999; 99US-0154336.  
XX (INCY-) INCYTE PHARM INC.  
PA  
XX  
XX Tang YT, Yue H, Hillman JL, Cuegler KJ, Corley NC, Lal P;  
PI Azlmzai Y, Baughn MR, Yang J, Shih LL;  
XX  
XX WPI; 2000-339688/29.  
XX N-PSDB; AAA15006.  
XX  
XX New human proliferation and apoptosis related protein polypeptides used  
PT for diagnosis, treatment and prevention of cell proliferative,  
PT immunological and reproductive disorders -  
XX  
XX Claim 1; Page 106-107; 128pp; English.  
XX  
XX The present sequence represents a human proliferation and apoptosis  
CC related protein (PROAP). The polypeptides and polynucleotides can be  
CC used for the diagnosis, treatment and prevention of cell proliferative,  
CC immunological and reproductive disorders. Disorders associated with  
CC decreased expression or activity of include arteriosclerosis, cirrhosis,  
CC hepatitis, psoriasis, melanoma, lymphoma and cancers of the breast,  
CC brain and prostate, acquired immune deficiency syndrome (AIDS),  
CC allergies, anaemia, asthma, diabetes mellitus, osteoarthritis,  
CC endometriosis, uterine fibroids and disruptions of the menstrual cycle.  
CC Antibodies against PROAP can be used in diagnosis of disorders  
CC characterized by PROAP e.g. in ELISA (enzyme linked immunosorbent  
CC assays) and the polynucleotides may be used to detect and quantify gene  
CC expression in biopsied tissues. These techniques can also be used to  
CC monitor regulation of PROAP levels during therapeutic intervention.  
XX  
XX Sequence 526 AA;  
SQ  
  
Query Match 48.5%; Score 16; DB 21; Length 526;  
Best Local Similarity 17.6%; Pred. No. 1.4e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
Qy 2 GXXXXXXXGXXXXXXG 18  
||  
Db 23 GSSSGGTTTTTTTGG 39  
  
RESULT 8  
AAB94371  
ID AAB94371 standard; Protein; 526 AA.  
XX  
XX AAB94371;  
AC  
XX  
XX 26-JUN-2001 (first entry)  
DT  
XX  
XX Human protein sequence SEQ ID NO:14909.  
DE  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX EP1074617-A2.  
PN  
XX  
XX 07-FEB-2001.  
PD  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
PF  
XX  
XX 29-JUL-1999; 99JP-0248036.  
PR  
XX  
XX 27-AUG-1999; 99JP-0300253.  
PR  
XX  
XX 11-JAN-2000; 2000JP-0118776.  
PR  
XX  
XX 02-MAY-2000; 2000JP-0183767.  
PR  
XX  
XX 09-JUN-2000; 2000JP-0241899.  
PR  
XX  
XX (HELI-) HELIX RES INST.  
PA  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
PI

XX WPI; 2001-318749/34.  
DR  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 14909; 2537pp + CD ROM; English.  
PS  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX Sequence 526 AA;  
SQ  
  
Query Match 48.5%; Score 16; DB 22; Length 526;  
Best Local Similarity 17.6%; Pred. No. 1.4e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
Qy 2 GXXXXXXXGXXXXXXG 18  
||  
Db 23 GSSSGGTTTTTTTGG 39  
  
RESULT 9  
AAY02540  
ID AAY02540 standard; Protein; 623 AA.  
XX  
XX AAY02540;  
AC  
XX  
XX 16-JUL-1999 (first entry)  
DT  
XX  
XX Protein encoded by wheat Rht clone 5a1 genomic sequence.  
DE  
XX  
XX Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
KW paclobitrazol.  
KW  
XX  
XX Triticum aestivum.  
OS  
XX  
XX WO9909174-A1.  
PN  
XX  
XX 25-FEB-1999.  
PD  
XX  
XX 07-AUG-1998; 98WO-GB02383.  
PF  
XX  
XX 13-AUG-1997; 97GB-0017192.  
PR  
XX  
XX (PLAN-) PLANT BIOSCIENCE LTD.  
PA  
XX  
XX Harberd NP, Peng J, Richards DE;  
XX

Wed Feb 13 07:51:52 2002

us-09-485-571-20.rag

DR WPI; 1999-181040/15.  
 DR N-PSDB; AAX36279.  
 XX New Triticum aestivum polynucleotides - encode a polypeptide which  
 PT provides inhibition of the growth of plants, which inhibition is  
 PT antagonised by gibberellin, used to confer a dwarf phenotype  
 XX  
 PS Disclosure; Fig 8b; 88pp; English.  
 XX  
 CC The specification describes polypeptides encoded by the Rht gene (and  
 CC its homologues) that, when expressed in Triticum aestivum, inhibit  
 CC growth of the plant. This growth inhibition is antagonised by  
 CC gibberellin. The products can be used to provide Rht expression in  
 CC plants, conferring a dwarf phenotype on a plant which is correctable  
 CC by treatment with gibberellin. In addition, the products can be  
 CC used to produce Rht mutant plants which are dwarfed compared with  
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
 CC may be made by knocking out Rht or the relevant homologous gene in  
 CC the plant of interest. Plants may be made which are resistant to  
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
 CC dwarf but let crop plants grow tall. The present sequence is encoded by  
 CC the wheat Rht clone 5a1 genomic sequence.  
 XX  
 SQ Sequence 623 AA;  
 Query Match 48.5%; Score 16; DB 20; Length 623;  
 Best Local Similarity 17.6%; Pred. No. 1.5e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 GGXXXXXXXXXXXXXXG 18  
 ||  
 Db 181 ggsstssssssslgg 197  
 RESULT 10  
 AAR13949  
 ID AAR13949 standard; Protein: 736 AA.  
 XX  
 AC AAR13949;  
 XX  
 DT 28-NOV-1991 (first entry)  
 DE SUP-B27 t(1;19) translocation fusion protein - clone KJ9.  
 XX  
 DE Immunoglobulin; enhancer-binding factor; E12; E47; E2A; homeoprotein;  
 KW chromosomal translocation; leukemia; fusion protein.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..477 "E2A identical sequence"  
 FT Protein /note="E2A identical sequence"  
 FT Protein 478..736  
 FT Region /note="chromosome 1-derived sequence"  
 FT Protein 621..684  
 FT Protein /label="homeodomain"  
 XX  
 XX WO9113172-A.  
 PN  
 XX  
 XX 05-SEP-1991.  
 PD  
 XX 22-FEB-1991; 91WO-US01168.  
 PF  
 XX 23-FEB-1990; 90US-0484063.  
 PR  
 XX (STRD ) LELAND STANFORD JR UNIV.  
 PA (WHIT-) WHITEHEAD INST BIOMED RE.  
 XX  
 XX Cleary ML, Mellentin JD, Baltimore D, Murre C, Mccaw P;  
 PI WPI; 1991-281484/38.  
 XX  
 DR

DR N-PSDB; AAQ13673.  
 XX  
 CC Detection of t(1;19) break-point-associated genes E2A and pr1 -  
 PT in chromosomal translocation, and prods. useful in diagnosis and  
 PT therapy of human neoplasm, esp. acute lymphoblastic leukaemia  
 XX  
 PS Disclosure; Fig 8; 104pp; English.  
 XX  
 CC The E2A gene (see AAQ13669), encoding immunoglobulin enhancer-binding  
 CC factors E12 and E47 (see AAQ13670 and AAQ13671), is localised to the  
 CC breakpoint of a consistently recurring chromosomal translocation  
 CC present in many acute leukemias and is structurally altered by most  
 CC t(1;19) chromosomal translocations. The translocation results in  
 CC synthesis of a fusion mRNA (see AAQ13672 for fusion cDNA) that crosses  
 CC the breakpoint between chromosome 1 and 19.  
 CC The complete nucleotide sequences of the SUP-B27 fusion cDNAs were  
 CC determined and represented in a composite sequence (see AAQ13672).  
 CC Clone KJ9 differs from the other four fusion cDNAs in that it has a  
 CC small deletion of chromosome 1-derived DNA. The KJ9 variant  
 CC encodes a predicted 80 kD protein.  
 CC Although the predicted fusion proteins contd. the amino two-thirds of  
 CC E2A, they no longer retained the helix-loop-helix DNA-binding and  
 CC dimerisation motif, which was replaced by a polypeptide encoded by  
 CC DNA from chromosome 1. The indicated region (amino acids 621-684)  
 CC overlaps with a region in homeoproteins that correspond to their  
 CC homeodomains.  
 CC See also AAQ13669-75.  
 XX  
 SQ Sequence 736 AA;  
 Query Match 48.5%; Score 16; DB 12; Length 736;  
 Best Local Similarity 17.6%; Pred. No. 1.7e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 GCXXXXXXXXXXXXXXG 18  
 ||  
 Db 513 ggsaaaaaaaaggag 529  
 RESULT 11  
 AAR15158  
 ID AAR15158 standard; Protein: 742 AA.  
 XX  
 AC AAR15158;  
 XX  
 DT 28-NOV-1991 (first entry)  
 DE E2A/pr1 fusion protein TYPE II from clone 697-4.  
 XX  
 DE Immunoglobulin; enhancer-binding factor; chromosomal translocation;  
 KW leukemia; fusion protein.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..484  
 FT Protein /label="E2A"  
 FT Protein 484..485  
 FT Protein /label="E2A/pr1\_breakpoint"  
 FT Protein 485..742  
 FT Protein /label="pr1"  
 XX  
 XX WO9113172-A.  
 PN  
 XX  
 XX 05-SEP-1991.  
 PD  
 XX 22-FEB-1991; 91WO-US01168.  
 PF  
 XX 23-FEB-1990; 90US-0484063.  
 PR  
 XX (STRD ) LELAND STANFORD JR UNIV.  
 PA (WHIT-) WHITEHEAD INST BIOMED RE.  
 XX  
 XX

XX Cleary ML, Mellentin JD, Baltimore D, Murre C, Mccaw P;  
 PI WPI: 1991-281484/38.  
 DR N-PSDB; AAQ13674.  
 XX  
 PT Detection of t(1;19) break-point-associated genes E2A and prl -  
 PT in chromosomal translocation, and prods. useful in diagnosis and  
 PT therapy of human neoplasm, esp. acute lymphoblastic leukaemia  
 XX  
 PS Disclosure; Fig 4B; 104pp; English.  
 XX  
 CC Fusion protein Type I is represented in AAR13951.  
 CC See also AAQ13669-75.  
 XX  
 SQ Sequence 742 AA;  
 Query Match 48.5%; Score 16; DB 12; Length 742;  
 Best Local Similarity 17.6%; Pred. NO. 1.7e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 GGXXXXXXXG 18  
 Db 519 ggsaaaaaaaggag 535  
 RESULT 12  
 AAR13948  
 ID AAR13948 standard; Protein; 819 AA.  
 XX  
 AC AAR13948;  
 XX  
 XX 28-NOV-1991 (first entry)  
 DE SUP-B27 t(1;19) translocation fusion protein.  
 XX  
 XX Immunoglobulin; enhancer-binding factor; E12; E47; E2A; homeoprotein;  
 KW chromosomal translocation; leukemia; fusion protein.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..477  
 FT /note= "E2A identical sequence"  
 FT 478..819  
 FT /note= "chromosome 1-derived sequence"  
 FT 621..684  
 FT /label= homeodomain  
 XX  
 PN WO9113172-A.  
 XX  
 PD 05-SEP-1991.  
 XX  
 PF 22-FEB-1991; 91WO-US01168.  
 XX  
 PR 23-FEB-1990; 90US-0484063.  
 XX  
 PA (STRD ) LELAND STANFORD JR UNIV.  
 PA (WHIT-) WHITEHEAD INST BIOMED RE.  
 XX  
 XX Cleary ML, Mellentin JD, Baltimore D, Murre C, Mccaw P;  
 XX WPI: 1991-281484/38.  
 DR N-PSDB; AAQ13672.  
 XX  
 XX Detection of t(1;19) break-point-associated genes E2A and prl -  
 PT in chromosomal translocation, and prods. useful in diagnosis and  
 PT therapy of human neoplasm, esp. acute lymphoblastic leukaemia  
 XX  
 PS Disclosure; Fig 8; 104pp; English.  
 XX  
 CC The E2A gene (see AAQ13669), encoding immunoglobulin enhancer-binding

CC factors E12 and E47 (see AAQ13670 and AAQ13671), is localised to the  
 CC breakpoint of a consistently recurring chromosomal translocation  
 CC present in many acute leukemias and is structurally altered by most  
 CC t(1;19) chromosomal translocations. The translocation results in  
 CC synthesis of a fusion mRNA (see AAQ13672 for fusion cDNA) that crosses  
 CC the breakpoint between chromosome 1 and 19.  
 CC The complete nucleotide sequences of the SUP-B27 fusion cDNAs were  
 CC determined and represented in a composite sequence (AAQ13672).  
 CC The fusion cDNAs encode a 85 kD protein. Clone KJ9 differs from the  
 CC other four fusion cDNAs in that it has a small deletion of chromosome  
 CC 1-derived DNA (see AAQ13673).  
 CC Although the predicted fusion proteins contd. the amino two-thirds of  
 CC E2A, they no longer retained the helix-loop-helix DNA-binding and  
 CC dimerisation motif, which was replaced by a polypeptide encoded by  
 CC DNA from chromosome 1. The indicated region (amino acids 621-684)  
 CC overlaps with a region in homeoproteins that correspond to their  
 CC homeodomains.  
 CC See also AAQ13669-75.  
 XX  
 SQ Sequence 819 AA;  
 Query Match 48.5%; Score 16; DB 12; Length 819;  
 Best Local Similarity 17.6%; Pred. NO. 1.8e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 GGXXXXXXXG 18  
 Db 513 ggsaaaaaaaggag 529  
 RESULT 13  
 AAR13951  
 ID AAR13951 standard; Protein; 825 AA.  
 XX  
 AC AAR13951;  
 XX  
 XX 28-NOV-1991 (first entry)  
 DE E2A/prl fusion protein TYPE I from clone 697-4.  
 XX  
 KW Immunoglobulin; enhancer-binding factor; chromosomal translocation;  
 KW leukemia; fusion protein.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..484  
 FT /label= E2A  
 FT 484..485  
 FT /label= E2A/prl\_breakpoint  
 FT 485..825  
 FT /label= prl  
 XX  
 PN WO9113172-A.  
 XX  
 PD 05-SEP-1991.  
 XX  
 PF 22-FEB-1991; 91WO-US01168.  
 XX  
 PR 23-FEB-1990; 90US-0484063.  
 XX  
 PA (STRD ) LELAND STANFORD JR UNIV.  
 PA (WHIT-) WHITEHEAD INST BIOMED RE.  
 XX  
 XX Cleary ML, Mellentin JD, Baltimore D, Murre C, Mccaw P;  
 XX WPI: 1991-281484/38.  
 DR N-PSDB; AAQ13674.  
 XX  
 XX Detection of t(1;19) break-point-associated genes E2A and prl -  
 PT in chromosomal translocation, and prods. useful in diagnosis and  
 PT therapy of human neoplasm, esp. acute lymphoblastic leukaemia

XX PS  
XX CC  
XX CC  
XX CC  
XX SQ

Disclosure; Fig 4B; 104pp; English.  
Fusion protein Type II is represented in AAR15158.  
See also AAQ13669-75.  
Sequence 825 AA;

Query Match 48.5%; Score 16; DB 12; Length 825;  
Best Local Similarity 17.6%; Pred. No. 1.8e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 519 ggsaaaaaaaggag 535

RESULT 14  
AAG02803  
ID AAG02803 standard; Protein; 104 AA.

XX AC AAG02803;  
XX DT 06-OCT-2000 (first entry)  
XX DE Human secreted protein, SEQ ID NO: 6884.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST ) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX DR N-PSDB; AAC02809.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX PS Claim 13; SEQ ID 6884; 71pp + CD-ROM; English.

XX CC The present sequence is a polypeptide encoded by one of a large number  
XX CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
XX CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
XX CC different tissues. EST sequences usually correspond mainly to the 3'  
XX CC untranslated region (UTR) of the mRNA because they are often obtained  
XX CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
XX CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
XX CC those cases where longer cDNA sequences have been obtained, the full 5'  
XX CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
XX CC ends and can therefore be used to obtain full length cDNAs and genomic  
XX CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
XX CC chromosome mapping procedures. They are used to obtain upstream  
XX CC regulatory sequences and to design expression and secretion vectors.

XX SQ Sequence 104 AA;

Query Match 45.5%; Score 15; DB 21; Length 104;  
Best Local Similarity 17.6%; Pred. No. 7.8e+03;

Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 5 99gsgsskassassag 21

RESULT 15  
AAV02538  
ID AAY02538 standard; Protein; 256 AA.

XX AC AAY02538;

XX DT 16-JUL-1999 (first entry)

XX DE Protein encoded by rice EST D39460 sequence.

XX KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;  
XX KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;  
XX KW paclobutrazol; rice; expressed sequence tag; EST.

XX OS Oryza sativa.

XX PN WO9909174-A1.

XX PD 25-FEB-1999.

XX PF 07-AUG-1998; 98WO-GB02383.

XX PR 13-AUG-1997; 97GB-0017192.

XX PA (PLAN-) PLANT BIOSCIENCE LTD.

XX PI Harberd NP, Peng J, Richards DE;

XX DR WPI; 1999-181040/15.

XX DR N-PSDB; AAX36277.

XX PT New Triticum Aestivum polynucleotides - encode a polypeptide which  
XX PT provides inhibition of the growth of plants, which inhibition is  
XX PT antagonised by gibberellin, used to confer a dwarf phenotype

XX PS Claim 12; Fig 6b; 88pp; English.

XX CC The specification describes polypeptides encoded by the Rht gene (and  
XX CC its homologues) that, when expressed in Triticum Aestivum, inhibit  
XX CC growth of the plant. This growth inhibition is antagonised by  
XX CC gibberellin. The products can be used to provide Rht expression in  
XX CC plants, conferring a dwarf phenotype on a plant which is correctable  
XX CC by treatment with gibberellin. In addition, the products can be  
XX CC used to produce Rht mutant plants which are dwarfed compared with  
XX CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants  
XX CC may be made by knocking out Rht or the relevant homologous gene in  
XX CC the plant of interest. Plants may be made which are resistant to  
XX CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,  
XX CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds  
XX CC dwarf but let crop plants grow tall. The present sequence is encoded by  
XX CC rice expressed sequence tag (EST) AAD39460, which is homologous to the  
XX CC wheat Rht gene.

XX SQ Sequence 256 AA;

Query Match 45.5%; Score 15; DB 20; Length 256;  
Best Local Similarity 17.6%; Pred. No. 1.4e+04;

Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 184 gggstssssssslgg 200

Search completed: February 12, 2002, 12:30:31  
Job time: 364 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:24 ; Search time 106.12 Seconds  
(without alignments)  
3.605 Million cell updates/sec

Title: US-09-485-571-31

Perfect score: 22

Sequence: 1 XXXXXXXXXXXXXXXXXX 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	27.3	2	1	US-07-791-213D-39
2	6	27.3	2	1	US-07-729-353-5
3	6	27.3	2	1	US-08-122-510-9
4	6	27.3	2	1	US-08-122-510-14
5	6	27.3	2	1	US-08-191-866D-69
6	6	27.3	2	1	US-08-293-150A-39
7	6	27.3	2	1	US-08-470-837-6
8	6	27.3	2	2	US-08-272-255-2
9	6	27.3	2	2	US-08-185-949B-69
10	6	27.3	2	2	US-08-483-236-4
11	6	27.3	2	2	US-09-060-455-1
12	6	27.3	2	2	US-09-039-308A-4
13	6	27.3	2	4	US-08-789-333F-102
14	6	27.3	2	5	PCT-US95-08565-2
15	6	27.3	3	1	US-07-663-413-7
16	6	27.3	3	1	US-07-663-413-10
17	6	27.3	3	1	US-07-780-790A-1
18	6	27.3	3	1	US-07-816-679A-9
19	6	27.3	3	1	US-07-791-213D-38
20	6	27.3	3	1	US-08-118-135A-9
21	6	27.3	3	1	US-07-947-035-5
22	6	27.3	3	1	US-08-055-530-7
23	6	27.3	3	1	US-08-055-530-10
24	6	27.3	3	1	US-08-079-812-31
25	6	27.3	3	1	US-08-122-510-3
26	6	27.3	3	1	US-08-122-510-7
27	6	27.3	3	1	US-08-122-510-8

Sequence 11, Appl  
Sequence 12, Appl  
Sequence 56, Appl  
Sequence 6, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 7, Appl  
Sequence 10, Appl  
Sequence 26, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 8, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 7, Appl  
Sequence 9, Appl

#### ALIGNMENTS

#### RESULT 1

US-07-791-213D-39  
; Sequence 39, Application US/07791213D  
; Patent No. 5409895  
; GENERAL INFORMATION:  
; APPLICANT: MORISHITA, Hideaki  
; APPLICANT: KANAMORI, Toshinori  
; APPLICANT: NOBUHARA, Masahiro  
; TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE  
; TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME  
; TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF  
; TITLE OF INVENTION: TREATING USING THE SAME  
; NUMBER OF SEQUENCES: 108  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07791,213D  
; FILING DATE: 13-NOV-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-306745  
; FILING DATE: 13-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 029650-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-07-791-213D-39

Query Match 27.3%; Score 6; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;

Wed Feb 13 07:52:34 2002

us-09-485-571-31.rai

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 2 G 2

RESULT 2  
US-07-729-353-5  
; Sequence 5, Application US/07729353  
; Patent No. 5464819  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Naoyoshi  
; TITLE OF INVENTION: A No. 5464819el Physiologically Active Peptide Having  
; TITLE OF INVENTION: Immunoregulatory Activities  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 301 N. Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/729,353  
; FILING DATE: 19910712  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-182714  
; FILING DATE: 12-JUL-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svensson, Leonard R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 216-273P  
; TELEPHONE: 703 241 1300  
; TELEFAX: 703 532 3407  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-729-353-5

Query Match 27.3%; Score 6; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 1 G 1

RESULT 3  
US-08-122-510-9  
; Sequence 9, Application US/08122510  
; Patent No. 5464821  
; GENERAL INFORMATION:  
; APPLICANT: AASMUL-OLSEN, Stig  
; APPLICANT: WIDMER, Fred  
; TITLE OF INVENTION: Small peptidic compounds useful for the  
; TITLE OF INVENTION: treatment of Glaucoma  
; NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff, Ltd.  
STREET: 10 S. Wacker Dr.  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/122,510  
FILING DATE: 24-SEP-1993  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 0532/91  
FILING DATE: 25-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK92/00095  
FILING DATE: 25-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Iwanicki, John P  
REGISTRATION NUMBER: 34,628  
REFERENCE/DOCKET NUMBER: 93,848  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910/221-5317  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1...2  
OTHER INFORMATION: /note= "Where X is Benzylester  
OTHER INFORMATION:  
US-08-122-510-9

Query Match 27.3%; Score 6; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 1 G 1

RESULT 4  
US-08-122-510-14  
; Sequence 14, Application US/08122510  
; Patent No. 5464821  
; GENERAL INFORMATION:  
; APPLICANT: AASMUL-OLSEN, Stig  
; APPLICANT: WIDMER, Fred  
; TITLE OF INVENTION: Small peptidic compounds useful for the  
; TITLE OF INVENTION: treatment of Glaucoma  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 S. Wacker Dr.  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/122,510  
FILING DATE: 24-SEP-1993  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 0532/91  
FILING DATE: 25-MAR-1991  
PRIOR APPLICATION DATA: PCT/DK92/00095  
FILING DATE: 25-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Iwanicki, John P.  
REGISTRATION NUMBER: 34,628  
REFERENCE/DOCKET NUMBER: 93,848  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1234  
TELEFAX: 312-715-1000  
TELEX: 910/221-5317  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..2  
OTHER INFORMATION: /note= "Boc-Gly Val-Obzl"  
US-08-122-510-14

Query Match 27.3%; Score 6; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
Db 1 G 1

RESULT 5  
US-08-191-866D-69  
; Sequence 69, Application US/08191866D  
; Patent No. 5783195  
; GENERAL INFORMATION:  
; APPLICANT: Cochran, Mark D.  
; TITLE OF INVENTION: Recombinant Infectious Bovine  
; TITLE OF INVENTION: Rhinotracheitis Virus S-IBR-052 And Uses Thereof  
; NUMBER OF SEQUENCES: 99  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John P. White  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/191,866D  
; FILING DATE: 4 February 1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.

REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
TELEX: 422523  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-191-866D-69

Query Match 27.3%; Score 6; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
Db 2 G 2

RESULT 6  
US-08-293-150A-39  
; Sequence 39, Application US/08293150A  
; Patent No. 5792629  
; GENERAL INFORMATION:  
; APPLICANT: MORISHITA, Hideaki  
; APPLICANT: KANAMORI, Toshinori  
; APPLICANT: NOBUHARA, Masahiro  
; TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE  
; TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME  
; TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF  
; TITLE OF INVENTION: TREATING USING THE SAME  
; NUMBER OF SEQUENCES: 110  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/293,150A  
; FILING DATE: 19-AUG-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/791,213  
; FILING DATE: 13-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-306745  
; FILING DATE: 13-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 029650-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-293-150A-39

Wed Feb 13 07:52:34 2002

```

Query Match      27.3%; Score 6; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10
Db 2 G 2

RESULT 7
US-08-470-837-6
; Sequence 6, Application US/08470837
; Patent No. 5800811
; GENERAL INFORMATION:
; APPLICANT: Nimni, Marcel E.
; APPLICANT: Hall, Frederick L.
; APPLICANT: Tuan, Tai-Lan
; APPLICANT: Wu, Lingtao
; TITLE OF INVENTION: Transforming Growth Factor B Fusion
; TITLE OF INVENTION: and
; TITLE OF INVENTION: Their Use in Wound Healing
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025-3395
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,837
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: 30630-1US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-470-837-6

```

```

Query Match      27.3%; Score 6; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10
Db 1 G 1

RESULT 8
US-08-272-255-2
; Sequence 2, Application US/08272255
; Patent No. 5824859
; GENERAL INFORMATION:
; APPLICANT: Cashmore, Anthony R.
; APPLICANT: Ahmad, Margaret

```

```

; APPLICANT: Lin, Chentao
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
; TITLE OF INVENTION: Using the Same
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859-ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/272,255
; FILING DATE: 08-JUL-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: UPN-1795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-272-255-2

Query Match      27.3%; Score 6; DB 2; Length 2;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10
Db 2 G 2

RESULT 9
US-08-185-949B-69
; Sequence 69, Application US/08185949B
; Patent No. 5874279
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran
; APPLICANT: Richard D. Macdonald
; TITLE OF INVENTION: Recombinant Infectious Bovine
; TITLE OF INVENTION: Rhinotracheitis Virus
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,949B
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.

```

REGISTRATION NUMBER: 678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 278-0525  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-185-949B-69

Query Match 27.3%; Score 6; DB 2; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
Db 2 G 2

RESULT 10  
US-08-483-236-4  
Sequence 4, Application US/08483236  
Patent No. 5939385  
GENERAL INFORMATION:  
APPLICANT: Labroo, Virender  
APPLICANT: Busby, Sharon  
TITLE OF INVENTION: Transglutaminase Cross-Linkable  
TITLE OF INVENTION: Polypeptides and Methods Relating Thereto  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,236  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E  
REGISTRATION NUMBER: 31-648  
REFERENCE/DOCKET NUMBER: 93-09c1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6673  
TELEFAX: 206-442-6678  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-483-236-4

Query Match 27.3%; Score 6; DB 2; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
Db 1 G 1

RESULT 11  
US-09-060-455-1  
Sequence 1, Application US/09060455  
Patent No. 5965118  
GENERAL INFORMATION:  
APPLICANT: Duncan, Ruth  
APPLICANT: Evagorou, Evagoras  
APPLICANT: Buckley, Robert G.  
APPLICANT: Gianasi, Elisabetta  
TITLE OF INVENTION: Polymer-Platinum  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Ave. Ste. 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/060,455  
FILING DATE: 14-APR-1998  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/044,743  
FILING DATE: 18-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mohr, Judy M  
REGISTRATION NUMBER: 38,563  
REFERENCE/DOCKET NUMBER: 0495-0005.30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650)3240880  
TELEFAX: 650 324-0960  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-060-455-1

Query Match 27.3%; Score 6; DB 2; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
Db 1 G 1

RESULT 12  
US-09-039-308A-4  
Sequence 4, Application US/09039308A  
Patent No. 6069129  
GENERAL INFORMATION:  
APPLICANT: Sandberg, Lawrence; Roos, Phillip;  
APPLICANT: Mitts, Thomas  
TITLE OF INVENTION: ELASTIN DERIVED COMPOSITION  
TITLE OF INVENTION: AND METHOD OF  
TITLE OF INVENTION: USING SAME  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: REED SMITH SHAW & MCCLAY, LLP  
STREET: PO Box 488  
CITY: Pittsburgh

us-09-485-571-31.ra1

Wed Feb 13 07:52:34 2002

STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 15230  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
COMPUTER: Compaq  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Word 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/039,308A  
FILING DATE: March 13, 1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Miller, Raymond A.  
REGISTRATION NUMBER: 42,891  
REFERENCE/DOCKET NUMBER: 97-489  
TELEPHONE: (412) 288-4192  
TELEFAX: (412) 288-3300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-039-308A-4

Query Match 27.3%; Score 6; DB 3; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 2 G 2

RESULT 13  
US-08-789-333F-102  
Sequence 102, Application US/08789333F  
Patent No. 6153380  
GENERAL INFORMATION:  
APPLICANT: No. 6153380an, Garry P  
APPLICANT: Rothenberg, S. M.  
TITLE OF INVENTION: METHODS FOR SCREENING FOR TRANSDOMINANT INTRACELLULAR  
FILE REFERENCE: A642601DUBRMSDS  
CURRENT APPLICATION NUMBER: US/08/789,333F  
CURRENT FILING DATE: 1997-01-23  
PRIOR APPLICATION NUMBER: 08/589,108  
PRIOR FILING DATE: 1996-01-23  
PRIOR APPLICATION NUMBER: 08/589,911  
PRIOR FILING DATE: 1996-01-23  
NUMBER OF SEQ ID NOS: 102  
SOFTWARE: PatentIn ver. 2.0  
SEQ ID NO 102  
LENGTH: 2  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: linker  
US-08-789-333F-102

Query Match 27.3%; Score 6; DB 4; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 1 G 1

RESULT 14  
PCT-US95-08565-2  
Sequence 2, Application PC/TUS9508565  
GENERAL INFORMATION:  
APPLICANT: Cashmore, Anthony R.  
APPLICANT: Ahmad, Margaret  
APPLICANT: Lin, Chentao  
TITLE OF INVENTION: Blue Light Photoreceptors and Methods of  
TITLE OF INVENTION: Using the Same  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris  
STREET: One Liberty place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08565  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/272,255  
FILING DATE: 08-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Leary Ph.D., Kathryn  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: UPN-1795  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-08565-2

Query Match 27.3%; Score 6; DB 5; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 2 G 2

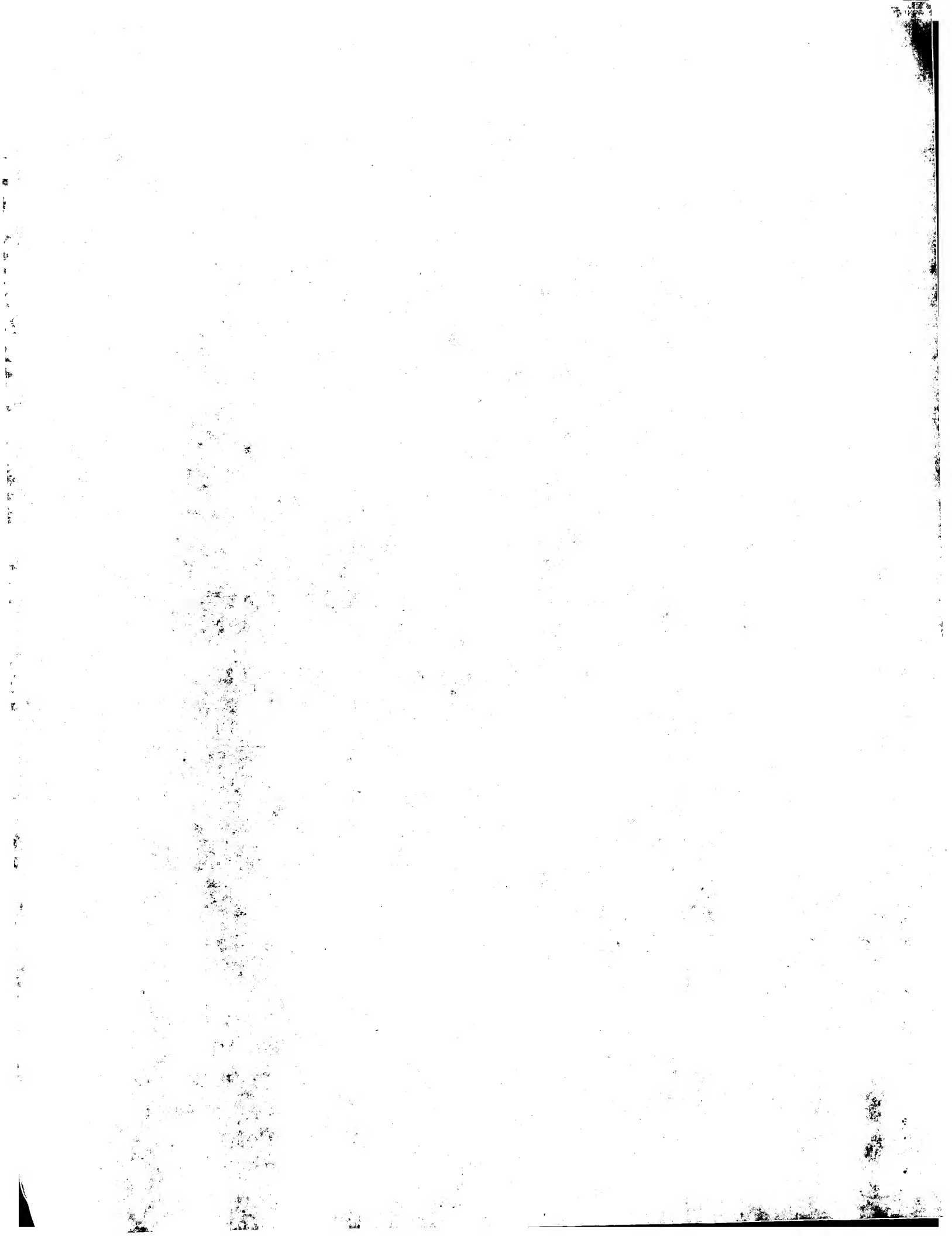
RESULT 15  
US-07-663-413-7  
Sequence 7, Application US/07663413  
Patent No. 5240703  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D.  
TITLE OF INVENTION: ATTENUATED, GENETICALLY-ENGINEERED  
TITLE OF INVENTION: PSEUDORABIES VIRUS S-PRV-155 AND USESTHEREOF  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White, Esq.  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/663,413  
FILING DATE: 19910301  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White Esq., John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 38720  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein,  
US-07-663-413-7

Query Match 27.3%; Score 6; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
|  
Db 3 G 3

Search completed: February 12, 2002, 12:32:25  
Job time: 453 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:42 ; Search time 126.85 Seconds  
(without alignments)  
10.209 Million cell updates/sec

Title: US-09-485-571-31

Perfect score: 22

Sequence: 1 XXXXXXXXXXXXXXXXXX 17

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	27.3	3	2	GKHU
2	6	27.3	3	2	A60898
3	6	27.3	3	2	A23751
4	6	27.3	3	2	B23751
5	6	27.3	3	2	PT0636
6	6	27.3	3	2	PT0571
7	6	27.3	3	1	ECXAA
8	6	27.3	4	2	A32039
9	6	27.3	4	2	PL0140
10	6	27.3	4	2	S09478
11	6	27.3	4	2	T30569
12	6	27.3	4	2	I38888
13	6	27.3	4	2	A25844
14	6	27.3	4	2	A34626
15	6	27.3	4	2	A32480
16	6	27.3	4	2	S39390
17	6	27.3	4	2	PT0240
18	6	27.3	4	2	PT0271
19	6	27.3	4	2	S43959
20	6	27.3	4	2	A53284
21	6	27.3	4	2	B53284
22	6	27.3	4	2	PT0633
23	6	27.3	4	2	PT0711
24	6	27.3	4	2	PT0698
25	6	27.3	4	2	PT0677
26	6	27.3	4	2	PT0706
27	6	27.3	4	2	PT0675
28	6	27.3	4	2	PT0721
29	6	27.3	4	2	PT0566

30	6	27.3	4	2	S47552
31	6	27.3	5	2	A32516
32	6	27.3	5	2	C23751
33	6	27.3	5	2	A26830
34	6	27.3	5	2	A41225
35	6	27.3	5	2	I40702
36	6	27.3	5	2	A44955
37	6	27.3	5	2	F22565
38	6	27.3	5	2	A33882
39	6	27.3	5	2	PQ0889
40	6	27.3	5	2	S51077
41	6	27.3	5	2	B45525
42	6	27.3	5	2	S65726
43	6	27.3	5	2	B61445
44	6	27.3	5	2	A61445
45	6	27.3	5	2	S11075

#### ALIGNMENTS

RESULT 1

GKHU

growth-modulating peptide - human

C:Species: Homo sapiens (man)

C:Date: 23-Oct-1981 #sequence\_revision 26-May-1995 #text\_change 20-Jun-2000

C:Accession: A01421

R:Schlesinger, D.H.; Pickart, L.; Thaler, M.M.

Experientia 33, 324-325, 1977

A:Title: Growth-modulating serum tripeptide is glycyl-histidyl-lysine.

A:Reference number: A01421; MUID:77162369

A:Accession: A01421

A:Molecule type: protein

A:Residues: 1-3 <SCH>

A:Note: this serum tripeptide is found to stimulate growth of some cell types and

C:Superfamily: unassigned animal peptides

Query Match 27.3% Score 6; DB 2; Length 3;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10

Db 1 G 1

RESULT 2

A60898

bursin - chicken

C:Species: Gallus gallus (chicken)

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 18-Aug-2000

C:Accession: A60898

R:Audhya, T.; Kroon, D.; Heavner, G.; Viamontes, G.; Goldstein, G.

Science 231, 997-999, 1986

A:Title: Tripeptide structure of bursin, a selective B-cell-differentiating hormone

A:Reference number: A60898; MUID:86122916

A:Accession: A60898

A:Molecule type: protein

A:Residues: 1-3 <AUD>

C:Superfamily: unassigned animal peptides

C:Keywords: amidated carboxyl end; hormone

F;3/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 27.3% Score 6; DB 2; Length 3;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10

Db 3 G 3

QY	10	G	10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
----	----	---	----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

A:Molecule type: protein  
A:Residues: 1-4 <HOR>  
A:Experimental source: brain  
C:Superfamily: unassigned animal peptides  
C:Keywords: amidated carboxyl end  
F:4/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
|  
Db 4 G 4

RESULT 9  
PL0140  
carbon-monoxide dehydrogenase (EC 1.2.99.2) large chain - Pseudomonas carboxydohydrogena  
C:Species: Pseudomonas carboxydohydrogena  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 28-Apr-1993  
C:Accession: PL0140  
R:Kraut, M.; Hugendieck, I.; Herwig, S.; Meyer, O.  
Arch. Microbiol. 152, 335-341, 1989  
A:Title: Homology and distribution of CO dehydrogenase structural genes in carboxydotrop  
A:Reference number: PL0138; MUID:90055678  
A:Accession: PL0140  
A:Molecule type: protein  
A:Residues: 1-4 <RRR>  
C:Comment: Carbon-monoxide dehydrogenase consists of three polypeptide chains: large, me  
C:Keywords: Oxidoreductase

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
|  
Db 2 G 2

RESULT 10  
S09478  
globulin IV alpha subunit gamma chain, seed - cucurbit (fragment)  
N:Alternate names: LIS globulin alpha subunit gamma chain  
C:Species: Cucurbita sp. (Cucurbit)  
C:Date: 21-Nov-1993 #sequence\_revision 08-Nov-1996 #text\_change 08-Nov-1996  
C:Accession: S09478  
R:Ohmiya, M.; Hara, I.; Matsubara, H.  
Plant Cell Physiol. 21, 157-167, 1980  
A:Title: Pumpkin (Cucurbita sp.) seed globulin IV. Terminal sequences of the acidic and  
A:Reference number: S09066  
A:Accession: S09478  
A:Molecule type: protein  
A:Residues: 1-4 <OHM>

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
|  
Db 3 G 3

RESULT 11  
T30569  
hypothetical protein - Emericella nidulans  
C:Species: Emericella nidulans, Aspergillus nidulans  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T30569  
R:Morrice, J.; MacKenzie, D.A.; Parr, A.J.; Archer, D.B.  
Curr. Genet. 34, 379-385, 1998  
A:Title: Isolation and characterisation of the acetyl-CoA carboxylase gene from As  
A:Reference number: Z20869; MUID:99087906  
A:Accession: T30569  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-4 <MOR>  
A:Cross-references: EMBL:Y15996; NID:el285512; PID:el218041; PIDN:CAA75927.1

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
|  
Db 2 G 2

RESULT 12  
I38888  
COI intron 16 protein - Podospora anserina mitochondrion  
C:Species: mitochondrion Podospora anserina  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 07-Dec-1999  
C:Accession: I38888  
R:Cummings, D.J.; Michel, F.; McNally, K.L.  
Curr. Genet. 16, 381-406, 1989  
A:Title: DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit  
A:Reference number: A48327; MUID:90124722  
A:Accession: I38888  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-4 <CUM>  
A:Cross-references: GB:X55026; GB:M30937; GB:M61734  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC3  
C:Keywords: mitochondrion

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
|  
Db 3 G 3

RESULT 13  
A25844  
autho-RF amide neuropeptide - sea pansy (Renilla koellikeri)  
C:Species: Renilla koellikeri (Koelliker's sea pansy)  
C:Date: 21-May-1988 #sequence\_revision 30-Sep-1993 #text\_change 11-Jul-1997  
C:Accession: A25844  
R:Grimmelikhuijzen, C.J.P.; Groeger, A.  
FEBS Lett. 211, 105-108, 1987  
A:Title: Isolation of the neuropeptide pGlu-Gly-Arg-Phe-amide from the pennatulid  
A:Reference number: A25844  
A:Accession: A25844  
A:Molecule type: protein  
A:Residues: 1-4 <GRI>  
C:Keywords: amidated carboxyl end; neuropeptide; pyroglutamic acid  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F:4/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Wed Feb 13 07:52:35 2002

QY 10 G 10  
|  
Db 2 G 2

RESULT 14

A34626  
RCH-related neuropeptide - ferruginous spindle  
C:Species: Fuscus ferrugineus (ferruginous spindle)  
C:Date: 06-Jul-1990 #sequence\_revision 06-Jul-1990 #text\_change 31-Dec-1993  
C:Accession: A34626  
R:Kuroki, Y.; Kanda, T.; Kubota, I.; Fujisawa, Y.; Ikeda, T.; Miura, A.; Minamitake, Y.;  
Biochem. Biophys. Res. Commun. 167, 273-279, 1990  
A:Title: A molluscan neuropeptide related to the crustacean hormone, RCH.  
A:Reference number: A34626; MUID:90179762  
A:Accession: A34626  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-4 <KUR>  
C:Keywords: neuropeptide

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
|  
Db 3 G 3

RESULT 15

A32480  
achatin-I - giant African snail  
N:Contains: achatin-II  
C:Species: Achatina fulica (giant African snail)  
C:Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 17-Mar-1999  
C:Accession: A32480  
R:Kamatani, Y.; Minakata, H.; Kenny, P.T.M.; Iwashita, T.; Watanabe, K.; Funase, K.; Sun  
Biochem. Biophys. Res. Commun. 160, 1015-1020, 1989  
A:Title: Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina fulica feru  
A:Reference number: A32480; MUID:89273551  
A:Accession: A32480  
A:Molecule type: protein  
A:Residues: 1-4 <KAM>  
R:Ishida, T.; In, Y.; Inoue, M.; Yasuda-Kamatani, Y.; Minakata, H.; Iwashita, T.; Nomoto  
FEBS Lett. 307, 253-256, 1992  
A:Title: Effect of the D-Phe(2) residue on molecular conformation of an endogenous neuro  
(H-Gly-Phe-Ala-Asp-OH).  
A:Reference number: A44691; MUID:92354723  
A:Contents: annotation; X-ray crystallography, 0.85 angstroms  
A:Note: achatin-II has L-phenylalanine  
C:Keywords: D-amino acid  
F/2/Modified site: D-phenylalanine (Phe) #status experimental

Query Match 27.3%; Score 6; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
|  
Db 1 G 1

Search completed: February 12, 2002, 12:34:42  
Job time: 560 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:54 ; Search time 67.2 Seconds  
(without alignments)  
9.275 Million cell updates/sec

Title: US-09-485-571-31  
Perfect score: 22  
Sequence: 1 XXXXXXXXXXXXXXXX 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6	27.3	3	GRWM_HUMAN	P01157 homo sapien
2	6	27.3	4	ACHL_ACHFU	P35904 achatina fu
3	6	27.3	4	DCML_PSECH	P19916 pseudomonas
4	6	27.3	4	EOSI_HUMAN	P02731 homo sapien
5	6	27.3	5	AL14_CARMA	P81817 carcinus ma
6	6	27.3	5	PAP2_PARMA	P81864 pardachirus
7	6	27.3	5	TPIS_CANFA	P54714 canis famli
8	6	27.3	5	UF01_MOUSE	P38639 mus musculu
9	6	27.3	5	UXA4_CHLTR	P38005 chlamydia t
10	6	27.3	6	CIP1_MYTED	P13736 mytilus edu
11	6	27.3	6	CIP2_MYTED	P13737 mytilus edu
12	6	27.3	6	FARP_MONEX	P41966 moniezia ex
13	6	27.3	6	LOK1_LOCMI	P41491 locusta mig
14	6	27.3	7	ALL2_CARMA	P81805 carcinus ma
15	6	27.3	7	ALL3_CARMA	P81806 carcinus ma
16	6	27.3	7	ALL4_CARMA	P81807 carcinus ma
17	6	27.3	7	ALL5_CARMA	P81808 carcinus ma
18	6	27.3	7	ALL7_CVDPO	P82158 cydia pomon
19	6	27.3	7	FAR1_HELTI	P41871 heliosoma tr
20	6	27.3	7	FAR5_HIRME	P42564 hirudo medi
21	6	27.3	7	IGA0_DACDE	P06294 dactylium d
22	6	27.3	7	LANC_CARUI	P36960 carnobacter
23	6	27.3	7	MNPI_LEPDI	P42984 leptinotars
24	6	27.3	7	UF04_MOUSE	P38642 mus musculu
25	6	27.3	7	UH11_RAT	P56576 rattus norv
26	6	27.3	7	UN06_PINPS	P81675 pinus pinas
27	6	27.3	8	ACI_THUAL	P18691 thunnus alb
28	6	27.3	8	AKHG_GRYBI	P14086 gryllus bim
29	6	27.3	8	AKH_TABAT	P14595 tabanus atr
30	6	27.3	8	AL12_CARMA	P81815 carcinus ma
31	6	27.3	8	AL15_CARMA	P81818 carcinus ma
32	6	27.3	8	AL16_CARMA	P81819 carcinus ma
33	6	27.3	8	AL17_CARMA	P81820 carcinus ma

34	6	27.3	8	AL18_CARMA	P81821 carcinus ma
35	6	27.3	8	ALL1_CVDPO	P82152 cydia pomon
36	6	27.3	8	ALL3_CVDPO	P82154 cydia pomon
37	6	27.3	8	ALL4_CALVO	P41840 calliphora
38	6	27.3	8	ALL4_CVDPO	P82155 cydia pomon
39	6	27.3	8	ALL5_CALVO	P41841 calliphora
40	6	27.3	8	ALL5_CVDPO	P82156 cydia pomon
41	6	27.3	8	ALL6_CVDPO	P82157 cydia pomon
42	6	27.3	8	ALL7_CARMA	P81809 carcinus ma
43	6	27.3	8	ALL8_CARMA	P81811 carcinus ma
44	6	27.3	8	ALL9_CARMA	P81812 carcinus ma
45	6	27.3	8	CAD1_ENTFA	P13268 enterococcu

ALIGNMENTS

RESULT 1  
GRWM\_HUMAN STANDARD; PRT; 3 AA.  
AC P01157;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 21-JUL-1986 (Rel. 01, Last annotation update)  
DE GROWTH-MODULATING PEPTIDE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=77162369; PubMed=858356;  
RA Schlesinger D.H., Pickart L., Thaler M.M.;  
RT "Growth-modulating serum tripeptide is glycyl-histidyl-lysine.";  
RL Experientia 33:324-325(1977).  
CC -!- MISCELLANEOUS: THIS SERUM TRIPEPTIDE HAS BEEN FOUND TO STIMULATE  
CC GROWTH OF SOME CELL TYPES AND TO INHIBIT OTHER TYPES IN VITRO.  
DR PIR; A01421; GRHU.  
SQ SEQUENCE 3 AA; 340 MW; 6331E81000000000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
Db 1 G 1

RESULT 2  
ACHL\_ACHFU STANDARD; PRT; 4 AA.  
AC P35904;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ACHATIN-I.  
OS Achatina fulica (Giant African snail).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Stylommatophora;  
OC Achatinacea; Achatinidae; Achatina.  
OX NCBI\_TaxID=6530;  
RN [1]  
RP SEQUENCE, CHARACTERIZATION, AND SYNTHESIS.  
RC STRAIN=FERUSSAC; TISSUE=Ganglion;  
RX MEDLINE=89273551; PubMed=2597281;  
RA Kamatani Y., Minakata H., Kenny P.T.M., Iwashita T., Watanabe K.,  
RA Funase K., Sun X.P., Yongsiri A., Kim K.H., Novales-Li P.,  
RA Novales E.T., Kanapi C.G., Takeuchi H., Nomoto K.;  
RT "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina  
fulica Ferussac containing a D-amino acid residue.";  
RL Biochem. Biophys. Res. Commun. 160:1015-1020(1989).  
RN [2]

CHARACTERIZATION.

RC STRAIN=FERUSSAC; TISSUE=Heart atrium;

RX MEDLINE=91264856; PubMed=1675568;

RA Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K., Yoshida M., Harada A., Muneoka Y., Kobayashi M.;

RT "Purification of aachatin-I from the atria of the African giant snail, Achatina fulica, and its possible function."

RL Biochem. Biophys. Res. Commun. 177:847-853(1991).

RN [3]

RP X-RAY CRYSTALLOGRAPHY.

RX MEDLINE=93014529; PubMed=1399265;

RA Ishida T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H., Iwashita T., Nomoto K.;

RT "Crystal structure and molecular conformation of aachatin-I (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a D-amino acid residue."

RL Int. J. Pept. Protein Res. 39:258-264(1992).

CC -1- FUNCTION: NEUROEXCITATORY PEPTIDE; INCREASES THE IMPULSE FREQUENCY AND PRODUCES A SPIKE BROADENING OF THE IDENTIFIED HEART EXCITATORY NEURON (PON); ALSO ENHANCES THE AMPLITUDE AND FREQUENCY OF THE HEART BEAT. HAS ALSO AN EFFECT ON SEVERAL OTHER MUSCLES.

CC PIR; A32480; A32480.

KW Hormone; D-amino acid.

FT MOD\_RES 2 2 D-PHENYLALANINE.

SQ SEQUENCE 4 AA; 408 MW; 6AADD9C810000000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

Db 1 G 1

RESULT 3

ID DCML\_PSECH STANDARD; PRT; 4 AA.

AC P19916;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 01-DEC-1992 (Rel. 24, Last annotation update)

DE CARBON MONOXIDE DEHYDROGENASE LARGE CHAIN (EC 1.2.99.2) (FRAGMENT).

OS Pseudomonas carboxydohydrogena.

OC Bacteria; Proteobacteria.

OX NCBI\_TaxID=290;

RN [1]

RP SEQUENCE.

RX MEDLINE=90055678; PubMed=2818128;

RA Kraut M., Hugendieck I., Herwig S., Meyer O.;

RT "Homology and distribution of CO dehydrogenase structural genes in carboxydohydrophic bacteria."

RL Arch. Microbiol. 152:335-341(1989).

CC -1- CATALYTIC ACTIVITY: CO + H(2)O + ACCEPTOR = CO(2) + REDUCED ACCEPTOR.

CC -1- COFACTOR: MOLYBDENUM.

CC -1- SUBUNIT: CONSISTS OF THREE POLYPEPTIDE CHAINS: LARGE, MEDIUM, AND SMALL.

DR PIR; P10140; P10140.

KW Oxidoreductase; Molybdenum.

FT NON\_TER 4 4

SQ SEQUENCE 4 AA; 441 MW; 7761E876F000000000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

Db 2 G 2

RESULT 4

EOSI\_HUMAN

ID EOSI\_HUMAN STANDARD; PRT; 4 AA.

AC P02731.

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 21-JUL-1986 (Rel. 01, Last annotation update)

DE EOSINOPHILOTACTIC PEPTIDES.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE.

RX MEDLINE=76078412; PubMed=1060093;

RA Goetzl E.J., Austen K.F.;

RT "Purification and synthesis of eosinophilic tetrapeptides of human lung tissue: identification as eosinophil chemotactic factor of anaphylaxis."

RL Proc. Natl. Acad. Sci. U.S.A. 72:4123-4127(1975).

CC -1- MISCELLANEOUS: THESE PEPTIDES ARE RELEASED FROM MAST CELLS IN LUNG (AND OTHER TISSUES) DURING HYPERSENSITIVITY REACTIONS (ANAPHYLAXIS). THEIR ACTIVITIES, PREFERENTIALLY AFFECTING EOSINOPHILS, INCLUDE CHEMOTAXIS, CHEMOTACTIC DEACTIVATION, RELEASE OF ENZYMES, AND STIMULATION OF THE HEXOSE MONOPHOSPHATE SHUNT.

CC PIR; A03190; ETHUL.

CC VARIANT 1 1 V -> A (IN OTHER PEPTIDE).

FT /FTID=VAR\_005201.

SQ SEQUENCE 4 AA; 390 MW; 6B05B862A0000000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

Db 2 G 2

RESULT 5

AL14\_CARMA

ID AL14\_CARMA STANDARD; PRT; 5 AA.

AC P81817;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE CARCINUSTATIN 14.

OS Carcinus maenas (Common shore crab) (Green crab).

OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;

OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;

OC Eubranchyura; Portunioidea; Portunidae; Carcinus.

OX NCBI\_TaxID=6759;

RN [1]

RP SEQUENCE.

RX TISSUE=Cerebral ganglion, and Thoracic ganglion;

RX MEDLINE=98121193; PubMed=9461295;

RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P., Thorpe A.;

RT "Isolation and identification of multiple neuropeptides of the allatostatin superfamily in the shore crab Carcinus maenas."

RL Eur. J. Biochem. 250:727-734(1997).

CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.

CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.

KW Neuropeptide; Amidation; Multigene family.

FT MOD\_RES 5 5 AMIDATION (POTENTIAL).

SQ SEQUENCE 5 AA; 586 MW; 672879D5AB300000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 5;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

Db 4 G 4

## RESULT 6

PAP2\_PARMA STANDARD; PRT; 5 AA.

ID P81864;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PARDAXIN II (P11) (FRAGMENT)

OS Pardachirus marmoratus (Red sea moles sole).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;

OC Soleioidae; Soleidae; Pardachirus.

OX NCBI\_TaxID=31087;

RN [1]

RP SEQUENCE.

RC TISSUE=Skin secretion;

RX MEDLINE=87057369; PubMed=3782138;

RA Lazarovici P., Primor N., Loew L.M.;

RT "Purification and pore-forming activity of two hydrophobic

RT polypeptides from the secretion of the Red sea moles sole (Pardachirus

RT marmoratus).";

RL J. Biol. Chem. 261:16704-16713(1986).

CC -|- FUNCTION: EXHIBITS UNUSUAL SHARK REPELLENT AND SURFACTANT

CC PROPERTIES. FORMS VOLTAGE-DEPENDENT, ION-PERMEABLE CHANNELS

CC IN MEMBRANES. AT HIGH CONCENTRATION CAUSES CELL MEMBRANE LYSIS.

CC -|- SUBUNIT: MONOMER. IN AQUEOUS SOLUTION EXISTS AS A Tetramer.

CC -|- SUBCELLULAR LOCATION: SECRETED.

CC -|- SIMILARITY: BELONGS TO THE PARDAXIN FAMILY.

KW Toxin.

FT NON\_TER 5

SQ SEQUENCE 5 AA; 614 MW; 7769C9C8100000 CRC64;

## Query Match

Best Local Similarity 27.3%; Score 6; DB 1; Length 5;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

Db 1 G 1

## RESULT 7

TPIS\_CANFA

ID TPIS\_CANFA STANDARD; PRT; 5 AA.

AC P54714;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE TRIOSEPHOSPHATE ISOMERASE (EC 5.3.1.1) (TIM) (FRAGMENT).

GN TP11.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI\_TaxID=9615;

RN [1]

RP SEQUENCE.

RC TISSUE=Heart;

RX MEDLINE=98163340; PubMed=9504812;

RA Dunn M.J., Corbett J.M., Wheeler C.H.;

RT "HSC-2DPAGE and the two-dimensional gel electrophoresis database of

RT dog heart proteins.";

RL Electrophoresis 18:2795-2802(1997).

CC -|- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE = DIHYDROXY-

CC ACETONE PHOSPHATE.

CC -|- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.

CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -|- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.

DR HSC-2DPAGE; P54714; DOG.

DR InterPro: IPR000652; Trioseph\_isomrse.

DR PROSITE: PS00171; TIM; PARTIAL.

KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;

KW Pentose shunt.

FT NON\_TER 1

FT NON\_TER 5

SQ SEQUENCE 5 AA; 550 MW; 64444862C9A00000 CRC64;

## Query Match

Best Local Similarity 27.3%; Score 6; DB 1; Length 5;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

Db 3 G 3

## RESULT 8

UF01\_MOUSE

ID UF01\_MOUSE STANDARD; PRT; 5 AA.

AC P38639;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-FEB-1995 (Rel. 31, Last annotation update)

DE UNKNOWN PROTEIN FROM 2D-PAGE OF FIBROBLASTS (P19) (FRAGMENT).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE.

RC TISSUE=Fibroblast;

RX MEDLINE=95009907; PubMed=7523108;

RA Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;

RT "Separation and sequencing of familial and novel murine proteins

RT using preparative two-dimensional gel electrophoresis.";

RL Electrophoresis 15:735-745(1994).

CC -|- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN

CC PROTEIN IS: 6.6, ITS MW IS: 19 KDA.

FT NON\_TER 5

SQ SEQUENCE 5 AA; 717 MW; 7364087043100000 CRC64;

## Query Match

Best Local Similarity 27.3%; Score 6; DB 1; Length 5;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

Db 3 G 3

## RESULT 9

UXA4\_CHLTR

ID UXA4\_CHLTR STANDARD; PRT; 5 AA.

AC P38005;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE UNKNOWN PROTEIN FROM 2D-PAGE FROM ELEMENTARY BODY (FRAGMENT).

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=813;

RN [1]

RP SEQUENCE.

RC STRAIN=L2/434/BU;

RA Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,

RA Comanducci M., Christianen G., Birkelund S., Vretou E., Ratti G.,

RA Pallini V.;

RL Submitted (SEP-1994) to the SWISS-PROT data bank.  
 CC -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 4.5, ITS MW IS: 28 KDA.  
 FT NON\_TER 5  
 SQ SEQUENCE 5 AA; 474 MW; 75BAA865AA800000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
 Db 3 G 3

RESULT 10  
 CIP1\_MYTED STANDARD; PRT; 6 AA.  
 ID CIP1\_MYTED  
 AC P13736;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CONTRACTION-INHIBITING PEPTIDE I (MIP I).  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Pedal ganglion;  
 RX MEDLINE=88240357; PubMed=3377776;  
 RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;  
 RT "Structures and actions of Mytilus inhibitory peptides.";  
 RL Biochem. Biophys. Res. Commun. 152:1376-1382(1988).  
 CC -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN MUSCLES.  
 CC -!- SIMILARITY: TO MIP II.  
 DR PIR; A27696; A27696.  
 KW Hormone; Amidation.  
 FT MOD\_RES 6 6  
 SQ SEQUENCE 6 AA; 637 MW; 72C9C68775B81000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
 Db 1 G 1

RESULT 11  
 CIP2\_MYTED STANDARD; PRT; 6 AA.  
 ID CIP2\_MYTED  
 AC P13737;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CONTRACTION-INHIBITING PEPTIDE II (MIP II).  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Pedal ganglion;  
 RX MEDLINE=88240357; PubMed=3377776;  
 RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;  
 RT "Structures and actions of Mytilus inhibitory peptides.";  
 RL Biochem. Biophys. Res. Commun. 152:1376-1382(1988).  
 CC -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN

CC MUSCLES.  
 CC -!- SIMILARITY: TO MIP I.  
 DR PIR; B27696; B27696.  
 KW Hormone; Amidation.  
 FT MOD\_RES 6 6  
 SQ SEQUENCE 6 AA; 621 MW; 72C9C6876DD81000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
 Db 1 G 1

RESULT 12  
 FARP\_MONEX STANDARD; PRT; 6 AA.  
 ID FARP\_MONEX  
 AC P41966;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FMRFAMIDE-LIKE NEUROPEPTIDE GNFRF-AMIDE.  
 OS Moniezia expansa (Sheep tapeworm)  
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian platyhelminths;  
 OC Rhabditophora; Eulcithophora; Revertospermata; Mediofusata;  
 OC Neodermata; Cestoda; Eucestoda; Cyclophyllidae; Anoplocephalidae;  
 OC Moniezia.  
 OX NCBI\_TaxID=28841;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=93312289; PubMed=8323531;  
 RA Maule A.G., Shaw C., Halton D.W., Thim L.;  
 RT "GNFRFamide: a novel FMRFamide-immunoreactive peptide isolated from the sheep tapeworm, Moniezia expansa.";  
 RL Biochem. Biophys. Res. Commun. 193:1054-1060(1993).  
 CC -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE) FAMILY.  
 CC Neuropeptide; Amidation.  
 KW MOD\_RES 6 6  
 FT SEQUENCE 6 AA; 787 MW; 69D409C9C4481000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
 Db 1 G 1

RESULT 13  
 LOK1\_LOCM1 STANDARD; PRT; 6 AA.  
 ID LOK1\_LOCM1  
 AC P41491;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE LOCUSTAKININ I.  
 OS Locusta migratoria (Migratory locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
 OC Acridomorpha; Acridoidea; Acrididae; Locusta.  
 OX NCBI\_TaxID=7004;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Corpora cardiaca;  
 RX MEDLINE=92262851; PubMed=1585017;  
 RA Schoofs L., Holman G.M., Proost P., van Damme J., Hayes T.K.,  
 RA de Loof A.;



RT "Locustakinin, a novel myotropic peptide from Locusta migratoria,  
 RT isolation, primary structure and synthesis.";  
 RL Regul. Pept. 37:49-57(1992).  
 CC -1- FUNCTION: MYOTROPIC PEPTIDE. MAY BE IMPORTANT IN THE STIMULATION  
 CC OF ION TRANSPORT AND INHIBITION OF DIURETIC ACTIVITY IN MALPIGHIAN  
 CC TUBULES.  
 DR PIR: A61068; A61068.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 6 6  
 SQ SEQUENCE 6 AA; 654 MW; 686365A5B9CDB000 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
 |  
 Db 6 G 6

## RESULT 14

ID ALL2\_CARMA STANDARD; PRT; 7 AA.  
 AC P81805;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CARCINUSTATIN 2.  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Amidation; Multigene family.  
 FT MOD\_RES 7 7  
 SQ SEQUENCE 7 AA; 770 MW; 672879CDCB5DDB70 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
 |  
 Db 6 G 6

## RESULT 15

ID ALL3\_CARMA STANDARD; PRT; 7 AA.  
 AC P81806;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CARCINUSTATIN 3.  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]

RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Multigene family.  
 SQ SEQUENCE 7 AA; 796 MW; 672879CDCB476B70 CRC64;

Query Match 27.3%; Score 6; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
 |  
 Db 6 G 6

Search completed: February 12, 2002, 12:39:54  
 Job time: 807 sec

us-09-485-571-31.rsp

Wed Feb 13 07:52:36 2002

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:43 ; Search time 232.64 Seconds  
(without alignments)  
10.689 Million cell updates/sec

Title: US-09-485-571-31  
Perfect score: 22  
Sequence: 1 XXXXXXXXXXXXXXXX 17  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues 473505  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPREMBL\_17.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	27.3	5	13 P82073	P82073 litoria rub
2	6	27.3	7	2 Q47029	Q47029 enterobacte
3	6	27.3	7	2 Q47505	Q47505 escherichia
4	6	27.3	7	10 Q49223	Q49223 glycine max
5	6	27.3	7	10 P82445	P82445 nicotiana t
6	6	27.3	7	10 Q9C5B3	Q9C5B3 arabidopsis
7	6	27.3	7	11 Q63668	Q63668 rattus norv
8	6	27.3	7	11 Q63480	Q63480 rattus norv
9	6	27.3	7	11 Q55184	Q55184 rattus norv
10	6	27.3	7	12 Q07624	Q07624 rous sarcom
11	6	27.3	8	2 Q09258	Q09258 synechococc
12	6	27.3	8	2 Q56140	Q56140 streptococc
13	6	27.3	8	2 Q56429	Q56429 thermus aqu
14	6	27.3	8	2 Q85406	Q85406 coxiella bu
15	6	27.3	8	2 Q44463	Q44463 agrobacteri
16	6	27.3	8	2 Q9X3K1	Q9X3K1 prochloroco
17	6	27.3	8	2 Q9R5R2	Q9R5R2 shigella dy
18	6	27.3	8	2 Q9R4M3	Q9R4M3 enterococcu
19	6	27.3	8	2 Q9AGP4	Q9AGP4 arthrobacte

20	6	27.3	8	3 Q05403	Q05403 saccharomyc
21	6	27.3	8	3 Q9URB9	Q9URB9 saccharomyc
22	6	27.3	8	3 Q9HDS4	Q9HDS4 aspergillus
23	6	27.3	8	4 Q15889	Q15889 homo sapien
24	6	27.3	8	4 Q15901	Q15901 homo sapien
25	6	27.3	8	4 Q16468	Q16468 homo sapien
26	6	27.3	8	4 Q9Y4X6	Q9Y4X6 homo sapien
27	6	27.3	8	4 Q9UMC7	Q9UMC7 homo sapien
28	6	27.3	8	4 Q9UCN4	Q9UCN4 homo sapien
29	6	27.3	8	4 Q16428	Q16428 homo sapien
30	6	27.3	8	4 Q9P285	Q9P285 homo sapien
31	6	27.3	8	4 Q9P0K3	Q9P0K3 homo sapien
32	6	27.3	8	5 Q94695	Q94695 physarum po
33	6	27.3	8	5 Q02032	Q02032 lytechinus
34	6	27.3	8	5 Q94623	Q94623 manduca sex
35	6	27.3	8	5 Q9TWH6	Q9TWH6 perinerels
36	6	27.3	8	5 P82685	P82685 periplaneta
37	6	27.3	8	5 P82686	P82686 periplaneta
38	6	27.3	8	5 P82687	P82687 periplaneta
39	6	27.3	8	5 P82688	P82688 periplaneta
40	6	27.3	8	5 P82689	P82689 periplaneta
41	6	27.3	8	6 Q28866	Q28866 megaptera n
42	6	27.3	8	6 Q9XSX1	Q9XSX1 canis famil
43	6	27.3	8	6 Q9TT78	Q9TT78 canis famil
44	6	27.3	8	6 Q9TRY3	Q9TRY3 sus sp. ins
45	6	27.3	8	13 Q9PS69	Q9PS69 gallus gall

ALIGNMENTS

RESULT 1					
P82073					
ID P82073	PRELIMINARY;	PRT;	5 AA.		
AC P82073;					
DT 01-MAY-2000	(TrEMBLrel. 13, Created)				
DT 01-MAY-2000	(TrEMBLrel. 13, Last sequence update)				
DT 01-MAY-2000	(TrEMBLrel. 13, Last annotation update)				
DE RUBELLIDIN 3.2.					
OS Litoria rubella (Desert tree frog).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;					
OC Litoria.					
OX NCBI_caxID=104895;					
RN [1]					
RP SEQUENCE.					
RC TISSUE=SKIN SECRETION;					
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;					
RT "Peptides from the skin glands of the Australian buzzing tree frog					
RT Litori electrica. Comparison with the skin peptides from Litoria					
RT rubella.";					
RL Aust. J. Chem. 52:0-0(1999).					
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR					
CC ANTIBIOTIC ACTIVITY.					
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.					
KW Amphibian skin.					
SQ SEQUENCE 5 AA; 570 MW; 71A9C9C862A00000 CRC64;					
Query Match	27.3%;	Score 6;	DB 13;	Length 5;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 1;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy 10 G 10					
Db 2 G 2					
RESULT 2					
Q47029					
ID Q47029	PRELIMINARY;	PRT;	7 AA.		
AC Q47029;					
DT 01-NOV-1996	(TrEMBLrel. 01, Created)				

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
 OX NCBI\_TaxID=3847;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. ESSEX; TISSUE=ROOTS;  
 RX MEDLINE=91367679; PubMed=1891369;  
 RA Laux T., Goldberg R.B.;  
 RT "A plant DNA binding protein shares highly conserved sequence motifs  
 with HMg-box proteins";  
 RL Nucleic Acids Res. 19:4769-4769(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. ESSEX; TISSUE=ROOTS;  
 RA Mahalingam R., Knapp H.T.;  
 RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF047050; AAC03556.1;  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA; 850 MW; 6AAAAAB378637810 CRC64;

Query Match 27.3%; Score 6; DB 10; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 DB 1 G 1

RESULT 5

ID P82445 PRELIMINARY; PRT; 7 AA.  
 AC P82445;  
 DT 01-JUN-2000 (TrEMBLrel. 14, Created)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE 10 KDA CELL WALL PROTEIN (FRAGMENT).  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=CV. PETIT HAVANA;  
 RA Blee K.A., Bonham V.A., Mitchell G.P., Robertson D., Slabas A.R.,  
 RA Wojtaszek P., Bolwell G.P.;  
 RT "Proteomic study of secondary cell wall proteins from transformed  
 tobacco culture";  
 RL Planta 0:0-0(2000).  
 CC -!- SUBCELLULAR LOCATION: CELL WALL.  
 CC -!- TISSUE SPECIFICITY: XYLEM.  
 KW Cell wall.  
 FT NON\_TER 7  
 SQ SEQUENCE 7 AA; 758 MW; 69D2C1E862D1B2A0 CRC64;

Query Match 27.3%; Score 6; DB 10; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 DB 4 G 4

RESULT 6

ID Q9C5B3 PRELIMINARY; PRT; 7 AA.  
 AC Q9C5B3;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE AAD A1 PROTEIN (FRAGMENT).  
 GN AAD A1.

OS Enterobacter cloacae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Enterobacter.  
 OX NCBI\_TaxID=550;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=94079349; PubMed=82571126;  
 RA Rather P.N., Mann P.A., Mierzwia R., Hare R.S., Miller G.H., Shaw K.J.;  
 RT "Analysis of the aac(3)-Via gene encoding a novel 3'-N-  
 acetyltransferase";  
 RL Antimicrob. Agents Chemother. 37:2074-2079(1993).  
 DR EMBL; M88012; AAA16193.1;  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA; 744 MW; 633862D2C321A030 CRC64;

Query Match 27.3%; Score 6; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 DB 6 G 6

RESULT 3

ID Q47505 PRELIMINARY; PRT; 7 AA.  
 AC Q47505;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE PLASMID PMCC7 MCCA,B,C,D,E,F GENES.  
 GN MCCA.  
 OS Escherichia coli.  
 OG Plasmid PMCC7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96095297; PubMed=8522520;  
 RA Gonzalez-Pastor J.E., San Millan J.L., Castilla M.A., Moreno F.;  
 RT "Structure and organization of plasmid genes required to produce the  
 translation inhibitor microcin C7";  
 RL J. Bacteriol. 177:7131-7140(1995).  
 DR EMBL; X57583; CAA40808.1;  
 KW Plasmid.  
 SQ SEQUENCE 7 AA; 763 MW; 644DD44861B406F0 CRC64;

Query Match 27.3%; Score 6; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 DB 4 G 4

RESULT 4

ID Q49223 PRELIMINARY; PRT; 7 AA.  
 AC Q49223;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE HMG-I-LIKE PROTEIN (FRAGMENT).  
 OS Glycine max (Soybean).

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 GN HYPOTHETICAL 0.7 KDA PROTEIN (FRAGMENT).  
 OS DIDI 10A-2B.  
 GN Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ROOTS;  
 RX MEDLINE=21171025; PubMed=11277426;  
 RA Vercauteren I., Van Der Schueren E., Van Montagu M., Gheysen G.;  
 RT "Arabidopsis thaliana genes expressed in the early compatible  
 RT interaction with root-knot nematodes.";  
 RL Mol. Plant Microbe Interact. 14:288-299(2001).  
 DR EMBL: AJ286350; CAB71014.2; -;  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 FT 7 7  
 SQ SEQUENCE 7 AA; 719 MW; 6732C7287EB325D0 CRC64;

Query Match 27.3%; Score 6; DB 10; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 Db 4 G 4

RESULT 7  
 Q63668 PRELIMINARY; PRT; 7 AA.  
 ID Q63668  
 AC Q63668  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DE 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE VASOPRESSIN V2 RECEPTOR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;  
 RX MEDLINE=95396550; PubMed=7667072;  
 RA Mandon B., Bellanger A.C., Elalouf J.M.;  
 RT "Inverse PCR-mediated cloning of the promoter for the rat vasopressin  
 RT V2 receptor gene.";  
 RL Pflugers Arch. 430:12-18(1995).  
 DR EMBL: X83264; CAAS6237.1; -;  
 SQ SEQUENCE 7 AA; 703 MW; 75A767287DC6D6F0 CRC64;

Query Match 27.3%; Score 6; DB 11; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 Db 4 G 4

RESULT 8  
 Q63480 PRELIMINARY; PRT; 7 AA.  
 ID Q63480  
 AC Q63480  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)

DE TR4-NS ORPHAN RECEPTOR (FRAGMENT).  
 GN TR4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=96198747; PubMed=8612486;  
 RX Yoshikawa T., Makino S., Gao X.M., Xing G.Q., Chuang D.M.,  
 RA Detera-Wadleigh S.D.;  
 RT "Splice variants of rat TR4 orphan receptor: differential expression  
 RT of novel sequences in the 5'-untranslated region and C-terminal  
 RT domain.";  
 RL Endocrinology 137:1562-1571(1996).  
 DR EMBL: U59125; AAB02827.1; -;  
 FT NON\_TER 1 1  
 FT 7 7  
 SQ SEQUENCE 7 AA; 758 MW; 672AA87864005350 CRC64;

Query Match 27.3%; Score 6; DB 11; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 Db 4 G 4

RESULT 9  
 O55184 PRELIMINARY; PRT; 7 AA.  
 ID O55184  
 AC O55184  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE ORPHAN RECEPTOR TR4-NS (FRAGMENT).  
 GN TR4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=96198747; PubMed=8612486;  
 RX Yoshikawa T., Makino S., Gao X.M., Xing G.Q., Chuang D.M.,  
 RA Detera-Wadleigh S.D.;  
 RT "Splice variants of rat TR4 orphan receptor: differential expression  
 RT of novel sequences in the 5'-untranslated region and C-terminal  
 RT domain.";  
 RL Endocrinology 137:1562-1571(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96299786; PubMed=8661150;  
 RA Yoshikawa T., Dupont B.R., Leach R.J., Detera-Wadleigh S.D.;  
 RT "New variants of the human and rat nuclear hormone receptor, TR4:  
 RT expression and chromosomal localization of the human gene.";  
 RL Genomics 35:361-366(1996).  
 DR EMBL: U59454; AAB91433.1; -;  
 FT NON\_TER 1 1  
 FT 7 7  
 SQ SEQUENCE 7 AA; 663 MW; 6DDAA8787EB05350 CRC64;

Query Match 27.3%; Score 6; DB 11; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 G 10  
 Db 4 G 4

RESULT 10

```
OS Streptococcus thermophilus.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBL_TaxID=1308;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=ST11;
RX MEDLINE=95047254; PubMed=7958782;
RA Constable A., Mollet B.;
RT "Isolation and characterisation of promoter regions from Streptococcus
RL thermophilus.";
RL FEMS Microbiol. Lett. 122:85-90(1994).
DR EMBL; X78210; CAA55045.1; -
FT NON_TER 8
SQ SEQUENCE 8 AA; 846 MW; ED086772D5B045B6 CRC64;

Query Match 27.3%; Score 6; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10
DB 7 G 7

RESULT 13
ID Q56429 PRELIMINARY; PRT; 8 AA.
AC Q56429;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE GAPDH (FRAGMENT).
DE Thermus aquaticus (subsp. thermophilus).
OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX NCBL_TaxID=274;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=HB-8;
RX MEDLINE=89025722; PubMed=3052437;
RA Bowen D., Littlechild J.A., Fothergill J.E., Watson H.C., Hall L.;
RT "Nucleotide sequence of the phosphoglycerate kinase gene from the
RT extreme thermophile Thermus thermophilus. Comparison of the deduced
RT amino acid sequence with that of the mesophilic yeast phosphoglycerate
RT kinase.";
RL Biochem. J. 254:509-517(1988).
DR EMBL; X12464; CAA31005.1; -.
FT NON_TER 1
SQ SEQUENCE 8 AA; 885 MW; 33C873333732C72B CRC64;

Query Match 27.3%; Score 6; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10
DB 7 G 7

RESULT 14
ID O85406 PRELIMINARY; PRT; 8 AA.
AC O85406;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 1.0 KDA PROTEIN (FRAGMENT).
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Coxiella group; Coxiella.
OX NCBL_TaxID=777;

Q07624
ID Q07624 PRELIMINARY; PRT; 7 AA.
AC Q07624;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE HYPOTHETICAL PROTEIN (FRAGMENT).
OS Rous sarcoma virus.
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.
OX NCBL_TaxID=11886;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=PRAGUE C;
RX MEDLINE=93010967; PubMed=1327749;
RA Donze O., Spahr P.F.;
RT "Role of the open reading frames of Rous sarcoma virus leader RNA in
RT translation and genome packaging.";
RL EMBO J. 11:3747-3757(1992).
DR EMBL; X67587; CAA47862.1; -.
KW Hypothetical protein.
FT NON_TER 7
SQ SEQUENCE 7 AA; 672 MW; 776045A7687DD6F0 CRC64;

Query Match 27.3%; Score 6; DB 12; Length 7;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10
DB 3 G 3

RESULT 11
ID O09258 PRELIMINARY; PRT; 8 AA.
AC O09258;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-AUG-1998 (TReMBLrel. 07, Last annotation update)
DE NIFH (FRAGMENT).
CN NIFH.
OS Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).
OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.
OX NCBL_TaxID=41431;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=RF-1;
RA Chen H.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001780; AAC33369.1; -.
FT NON_TER 8
SQ SEQUENCE 8 AA; 985 MW; F16B59CDD046C406 CRC64;

Query Match 27.3%; Score 6; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10
DB 8 G 8

RESULT 12
ID Q56140 PRELIMINARY; PRT; 8 AA.
AC Q56140;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE (ST11) STP6 CONTROL REGION (FRAGMENT).
GN STP6.
```

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NINE MILE PHASE I;  
RA Williams H., Jaeger C., Baljer G.;  
RT "Physical and genetic map of the obligate intracellular bacterium  
RT Coxiella burnetii";  
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF064963; AAD09947.1; -  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 8 AA; 993 MW; 046BSAA453772727 CRC64;

Query Match 27.3%; Score 6; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

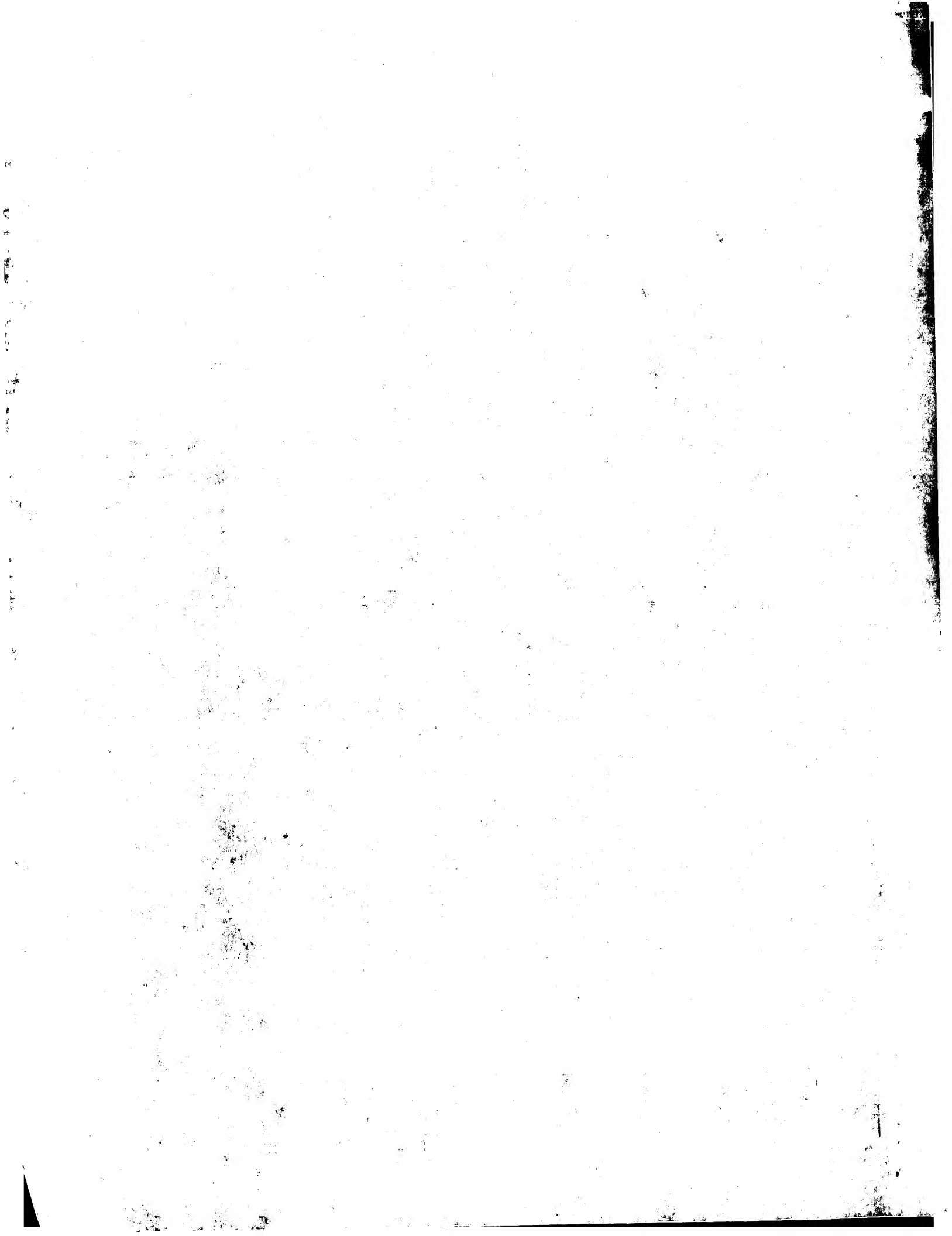
QY 10 G 10  
Db 8 G 8

RESULT 15  
Q44463  
ID Q44463 PRELIMINARY; PRT; 8 AA.  
AC Q44463;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE AGROBACTERIUM TUMEFACIENS TI PLASMID VIRID3 AND VIRID4 GENES (FRAGMENT).  
OS Agrobacterium radiobacter.  
OG Plasmid pTiA6NC.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=358;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88015611; PubMed=3658701;  
RA Portner S.G., Yanofsky M.F., Nester E.W.;  
RT "Molecular characterization of the virD operon from Agrobacterium  
RT tumefaciens";  
RL Nucleic Acids Res. 15:7503-7517(1987).  
DR EMBL; X06045; CAA29439.1; -  
KW Plasmid.  
FT NON\_TER 8  
SQ SEQUENCE 8 AA; 887 MW; F8F2C325B33861A6 CRC64;

Query Match 27.3%; Score 6; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 3 G 3

Search completed: February 12, 2002, 12:38:43  
Job time: 756 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:34 ; Search time 242.57 seconds  
(without alignments)  
5.191 Million cell updates/sec

Title: US-09-485-571-32

Perfect score: 90

Sequence: 1 KWAQFVAYGIRYLRL 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	100.0	17	20 AAW99416	Tachyplesin deriva
2	67	74.4	17	16 AAR75806	Antimicrobial tach
3	67	74.4	17	21 AAY69610	Tachyplesin analog
4	62	68.9	17	20 AAW99417	Tachyplesin deriva
5	61	67.8	17	16 AAR75808	Antimicrobial tach
6	61	67.8	17	20 AAW99413	Tachyplesin deriva
7	61	67.8	17	21 AAY93617	Peptide which may
8	60	66.7	17	16 AAR75819	Antimicrobial tach
9	59	65.6	17	16 AAR75807	Antimicrobial tach
10	59	65.6	17	21 AAY69609	Generic tachyplesi
11	59	65.6	21	16 AAR75816	Antimicrobial tach

12	59	65.6	35	16	AAW75810	Antimicrobial tach
13	58	64.4	17	16	AAW75822	Antimicrobial tach
14	58	64.4	17	20	AAW99414	Tachyplesin deriva
15	57	63.3	17	10	AAW91671	New lipopolysaccha
16	57	63.3	17	11	AAW06266	Antiviral peptide.
17	57	63.3	17	11	AAW06861	Tachyplesin I. Li
18	57	63.3	17	11	AAW08202	Gigaslin II. Tachy
19	57	63.3	17	13	AAW23112	Bacterial shock tr
20	57	63.3	17	13	AAW23114	Bacterial shock tr
21	57	63.3	17	14	AAW38491	Tachyplesin-III. T
22	57	63.3	17	14	AAW38489	Tachyplesin-I. Tac
23	57	63.3	17	16	AAW75805	Tachyplesin, an an
24	57	63.3	17	19	AAW66465	Cationic peptide t
25	57	63.3	17	21	AAW91764	Cationic peptide T
26	57	63.3	17	21	AAW69608	Tachyplesin (TP).
27	57	63.3	17	21	AAW69613	Tachyplesin analog
28	57	63.3	17	21	AAW69614	Tachyplesin analog
29	57	63.3	17	22	AAW91394	Tachyplesin peptid
30	57	63.3	39	16	AAW75817	Antimicrobial tach
31	56	62.2	17	16	AAW75820	Antimicrobial tach
32	56	62.2	17	21	AAW69611	Tachyplesin analog
33	55	61.1	17	16	AAW75803	Antimicrobial tach
34	55	61.1	17	21	AAW69612	Tachyplesin analog
35	54	60.0	17	11	AAW06862	Tachyplesin II. L
36	54	60.0	17	13	AAW23113	Bacterial shock tr
37	54	60.0	17	14	AAW38490	Tachyplesin-II. Ta
38	54	60.0	17	16	AAW75818	Antimicrobial tach
39	54	60.0	17	19	AAW66466	Cationic peptide t
40	54	60.0	17	21	AAW91765	Cationic peptide T
41	54	60.0	17	21	AAW69616	Tachyplesin analog
42	54	60.0	17	21	AAW69617	Tachyplesin analog
43	53	58.9	17	21	AAW69615	Tachyplesin analog
44	50	55.6	17	16	AAW75811	Antimicrobial tach
45	50	55.6	17	16	AAW75813	Antimicrobial tach

ALIGNMENTS

RESULT 1

AAW99416  
ID AAW99416 standard; peptide; 17 AA.

XX AC AAW99416;

XX DT 08-JUN-1999 (first entry)

XX DE Tachyplesin derivative peptide SM2191.

XX KW Linear: tachyplesin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

XX OS Synthetic.

XX PN WO990728-A2.

XX PD 18-FEB-1999.

XX PF 06-AUG-1998; 98WO-FR01757.

XX PR 12-AUG-1997; 97FR-0010297.

XX PA (SYNT-) SYNT:EM SA.

XX PI Calas E, Chavanieu A, Grassy G, Kaczorek M;

XX XX WPI; 1999-190034/16.

XX DR Derivatives of antibiotic peptides lacking disulfide bridges - used  
XX PT as carriers to deliver active agents into cells  
XX PT

Wed Feb 13 07:52:38 2002

us-09-485-571-32.rag

PS Claim 8; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the tachyplesin family of

CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide

CC antibiotic family which contain a beta-sheet secondary structure linked

CC by disulphide bridges. The new derivatives are linear and lack the

CC disulphide bridge. The novel derivatives are used to deliver active

CC agents to an organism, e.g. therapeutic proteins, antibodies (or their

CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,

CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic

CC and non-lytic but can cross mammalian cell membranes rapidly by a passive

CC mechanism, so can deliver active agents to cytoplasm and nucleus,

CC including crossing the blood-brain barrier.

XX Sequence 17 AA;

SQ

Query Match 100.0%; Score 90; DB 20; Length 17;

Best Local Similarity 100.0%; Pred. No. 4.8e-09;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KWAERVAYRGIRYLRL 17

Db 1 kwafrvayrgirylrl 17

RESULT 2

AA75806

ID AAR75806 standard; peptide; 17 AA.

AC AAR75806;

XX

XX

XX

XX

DT 07-FEB-1996 (first entry)

XX

DE Antimicrobial tachyplesin peptide derivative.

XX

XX Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;

KW infection.

KW

XX Synthetic.

OS

XX

XX WC9516776-AL.

PN

XX

XX 22-JUN-1995.

PD

XX

XX 19-DEC-1994; 94WO-US14619.

PF

XX

XX 17-DEC-1993; 93US-0168809.

PR

XX (PION-) PIONEER HI-BRED INT INC.

PA

XX Putman RJ, Rao AG, Rao A;

PI

XX WPI; 1995-231570/30.

DR

XX

XX New peptide derivs. of tachyplesin - having antimicrobial activity,

PT used against plant pathogenic fungi or human or animal infections

FT

XX

XX Claim 1; Page 29; 45pp; English.

PS

XX AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin

CC (AAR75805) a small peptide isolated from Japanese horseshoe crab

CC haemocytes which has antimicrobial properties. The peptide derivatives

CC also have antimicrobial activity and can be used for killing and

CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.

CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,

CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can

CC also be used for treating and preventing infection in humans and

CC animals.

XX

XX Sequence 17 AA;

SQ

Query Match 74.4%; Score 67; DB 16; Length 17;

Best Local Similarity 92.3%; Pred. No. 4.5e-05;

Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KWAERVAYRGIRY 13

Db 1 kwafrvayrgiray 13

RESULT 3

AA69610

ID AAY69610 standard; peptide; 17 AA.

XX

XX AAY69610;

AC

XX

XX 08-MAY-2000 (first entry)

DT

XX

XX Tachyplesin analogue, TPA.

DE

XX

XX Tachyplesin analogue; generic; antimicrobial; disulphide bond;

KW antifungal; antiviral; antimicrobial; transgenic plant.

KW

XX Synthetic.

OS

XX

XX Key Location/Qualifiers

PH Modified-site 17

FT /note= "C-terminal amide"

FT

XX

XX US6015941-A.

PN

XX

XX 18-JAN-2000.

PD

XX

XX 31-OCT-1997; 97US-0962034.

PF

XX

XX 31-OCT-1997; 97US-0962034.

PR

XX (PION-) PIONEER HI-BRED INT INC.

PA

XX Rao AG;

PI

XX

XX WPI; 2000-126327/11.

DR

XX

XX New tachyplesin analogs useful for controlling fungal and bacterial

PT activity in agricultural and medical applications and for controlling

PT plant viruses have four cysteine substitutions

PT

XX

XX Example 1; Page 17pp; English.

PS

XX Sequences AAY69610-Y69611 and AAY69617 represent tachyplesin analogues

CC used in an exemplification of the present invention, in which the

CC native tachyplesin cysteine residues are replaced with Ala, Leu and

CC Asp, respectively. Tachyplesin (AAY69608) is a naturally occurring

CC antimicrobial peptide which contains two disulphide bonds which help

CC to maintain its tertiary structure. The invention relates to novel

CC peptide analogues of tachyplesin (Y696912-AAY69614, AAY69616) in which

CC the cysteine residues at positions 3, 7, 12 and 16 of the native

CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,

CC valine, methionine, phenylalanine or tyrosine, the same amino acid being

CC present at all four positions. Despite being unable to form

CC intramolecular disulphide bonds, the analogues are functional as

CC antimicrobial agents. The tachyplesin analogues are useful for

CC controlling fungal and viral activity in agricultural and medical

CC applications and for controlling plant viruses. They can also be

CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,

CC soya or especially maize plants to provide resistance to pathogenic fungi

CC and viruses. Note: The present sequence is not shown in the

CC specification, but is derived from the generic tachyplesin analogue

CC sequence given in column 23.

XX

XX Sequence 17 AA;

SQ

Query Match 74.4%; Score 67; DB 21; Length 17;

```

Best Local Similarity 92.3%; Pred. No. 4.5e-05;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KWAFRVAYRGIRY 13
   |||||
Db 1 kwafrvayrgiay 13

RESULT 4
AAW99417
ID AAW99417 standard; peptide; 17 AA.
XX
AC AAW99417;
XX
DT 08-JUN-1999 (first entry)
XX
DE Tachyplesin derivative peptide SM2192.
XX
KW Linear; tachyplesin; peptide antibiotic; beta-sheet; secondary structure;
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;
KW nucleus; blood-brain barrier.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 17
FT /label= Nle

PN WO9907728-A2.
XX
PD 18-FEB-1999.
XX
PF 06-AUG-1998; 98WO-FR01757.
XX
PR 12-AUG-1997; 97FR-0010297.
XX
PA (SYNT-) SYNT:EM SA.
XX
PI Calas B, Chavanieu A, Grassy G, Kaczorek M;
XX
DR WPI; 1999-190034/16.
XX
PT Derivatives of antibiotic peptides lacking disulfide bridges - used
PT as carriers to deliver active agents into cells
XX
PS Claim 8; Page 28; 37pp; French.
XX
CC This peptide represents a linear derivative of the tachyplesin family of
CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide
CC antibiotic family which contain a beta-sheet secondary structure linked
CC by disulphide bridges. The new derivatives are linear and lack the
CC disulphide bridge. The novel derivatives are used to deliver active
CC agents to an organism, e.g. therapeutic proteins, antibodies (or their
CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,
CC antivirals, and anti-inflammatory, etc. The derivatives are non-toxic
CC and non-lytic but can cross mammalian cell membranes rapidly by a passive
CC mechanism, so can deliver active agents to cytoplasm and nucleus,
CC including crossing the blood-brain barrier.
XX
SQ Sequence 17 AA;

Query Match 68.9%; Score 62; DB 20; Length 17;
Best Local Similarity 75.0%; Pred. No. 0.00032;
Matches 12; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KWAFRVAYRGIRYLLR 16
   |||||
Db 1 kyawrvahrgirllr 16

RESULT 5

```

```

AAR75808
ID AAR75808 standard; peptide; 17 AA.
XX
AC AAR75808;
XX
DT 07-FEB-1996 (first entry)
XX
DE Antimicrobial tachyplesin peptide derivative.
XX
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;
KW infection.
XX
OS Synthetic.
XX
PN WO9516776-A1.
XX
PD 22-JUN-1995.
XX
PF 19-DEC-1994; 94WO-US14619.
XX
PR 17-DEC-1993; 93US-0168809.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Putman RJ, Rao AG, Rao A;
XX
DR WPI; 1995-231570/30.
XX
PT New peptide derivs. of tachyplesin - having antimicrobial activity,
PT used against plant pathogenic fungi or human or animal infections
XX
PS Claim 1; Page 30; 45pp; English.
XX
CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin
CC (AAR75805) a small peptide isolated from Japanese horseshoe crab
CC haemocytes which has antimicrobial properties. The peptide derivatives
CC also have antimicrobial activity and can be used for killing and
CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.
CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,
CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can
CC also be used for treating and preventing infection in humans and
CC animals.
XX
SQ Sequence 17 AA;

Query Match 67.8%; Score 61; DB 16; Length 17;
Best Local Similarity 76.9%; Pred. No. 0.00048;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KWAFRVAYRGIRY 13
   |||||
Db 1 kwlfvtyrgiky 13

RESULT 6
AAW99413
ID AAW99413 standard; peptide; 17 AA.
XX
AC AAW99413;
XX
DT 08-JUN-1999 (first entry)
XX
DE Tachyplesin derivative peptide SM1726.
XX
KW Linear; tachyplesin; peptide antibiotic; beta-sheet; secondary structure;
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;
KW nucleus; blood-brain barrier.
XX
OS Synthetic.
XX
PN WO9907728-A2.

```

XX PD 18-FEB-1999.  
XX PF 06-AUG-1998; 98WO-FR01757.  
XX PR 12-AUG-1997; 97FR-0010297.  
XX PA (SYNT-) SYNT:EM SA.  
XX PI Calas B, Chavanieu A, Grassy G, Kaczorek M;  
XX WPI; 1999-190034/16.  
XX DR Derivatives of antibiotic peptides lacking disulfide bridges - used  
XX PT as carriers to deliver active agents into cells  
XX PS Claim 8; Page 28; 37pp; French.  
XX CC This peptide represents a linear derivative of the tachyplesin family of  
XX CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide  
XX CC antibiotic family which contain a beta-sheet secondary structure linked  
XX CC by disulfide bridges. The new derivatives are linear and lack the  
XX CC disulfide bridge. The novel derivatives are used to deliver active  
XX CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
XX CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumor agents,  
XX CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
XX CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
XX CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
XX CC including crossing the blood-brain barrier.  
XX SQ Sequence 17 AA;

Query Match 67.8%; Score 61; DB 20; Length 17;  
Best Local Similarity 76.9%; Pred. NO. 0.00048;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KWAFFRVAYRGIRY 13  
   ||:|||||  
Db 1 kwsfrvsyrglsy 13

RESULT 7  
AAY93617  
ID AAY93617 standard; peptide; 17 AA.  
XX AC AAY93617;  
XX DT 25-SEP-2000 (first entry)  
XX DE Peptide which may be linked to anticancer agents.  
XX KW Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
XX KW cancer.  
XX OS Unidentified.  
XX PN WO200032237-A1.  
XX PD 08-JUN-2000.  
XX PF 26-NOV-1999; 99WO-FR02939.  
XX PR 30-NOV-1998; 98FR-0015073.  
XX PA (SYNT-) SYNT:EM SA.  
XX PI Tamsamani J, Kaczorek M, Colin De Verdiere A;  
XX WPI; 2000-412166/35.  
XX DR New composition useful for cancer treatment and prevention, contains  
XX PT anticancer agent and peptide vector that transports agent into cells

XX PS Disclosure; Page 8; 34pp; French.  
XX CC The specification describes a pharmaceutical composition, which  
XX CC comprises at least one anticancer agent associated with at least one  
XX CC peptide that can transport it into cancer cells and which inhibits  
XX CC development of resistance to the anticancer agent. By using the  
XX CC peptide as a vector for delivery of the anticancer agent, mechanisms  
XX CC that cause cancer cells to become resistant to the agent, particularly  
XX CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
XX CC produced by chemical synthesis, can be coupled easily to the agent,  
XX CC cross mammalian cell membranes rapidly by a passive mechanism (no  
XX CC receptors required), and are non-toxic and non-lytic. The compositions  
XX CC are used to treat cancer. The present sequence represents a peptide  
XX CC which may be linked to the anticancer agents of the invention.  
XX SQ Sequence 17 AA;

Query Match 67.8%; Score 61; DB 21; Length 17;  
Best Local Similarity 76.9%; Pred. NO. 0.00048;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KWAFFRVAYRGIRY 13  
   ||:|||||  
Db 1 kwsfrvsyrglsy 13

RESULT 8  
AAR75819  
ID AAR75819 standard; peptide; 17 AA.  
XX AC AAR75819;  
XX DT 07-FEB-1996 (first entry)  
XX DE Antimicrobial tachyplesin peptide derivative.  
XX KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
XX KW infection.  
XX OS Synthetic.  
XX PN WO9516776-A1.  
XX PD 22-JUN-1995.  
XX PF 19-DEC-1994; 94WO-US14619.  
XX PR 17-DEC-1993; 93US-0168809.  
XX PA (PION-) PIONEER HI-BRED INT INC.  
XX PI Putman RJ, Rao AG, Rao A;  
XX WPI; 1995-231570/30.  
XX DR New peptide derivs. of tachyplesin - having antimicrobial activity,  
XX PT used against plant pathogenic fungi or human or animal infections  
XX PS Claim 1; Page 35; 45pp; English.  
XX CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
XX CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
XX CC haemocytes which has antimicrobial properties. The peptide derivatives  
XX CC also have antimicrobial activity and can be used for killing and  
XX CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
XX CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
XX CC Sclerotinia trifoliorum and Aspergillus flauus. The peptides can  
XX CC also be used for treating and preventing infection in humans and  
XX CC animals.  
XX SQ Sequence 17 AA;

Query Match 66.7%; Score 60; DB 16; Length 17;  
 Best Local Similarity 76.9%; Pred. No. 0.00072;  
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KWAERVAYRGIRY 13  
 || ||| ||||:|  
 Db 1 kwfrvryrgiky 13

## RESULT 9

AAR75807  
 ID AAR75807 standard; peptide; 17 AA.

XX AC AAR75807;

DT 07-FEB-1996 (first entry)

XX Antimicrobial tachyplesin peptide derivative.

DE Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 XX infection.

XX Synthetic.

XX WO9516776-A1.

XX 22-JUN-1995.

XX 19-DEC-1994; 94WO-US14619.

XX 17-DEC-1993; 93US-0168809.

XX (PION-) PIONEER HI-BRED INT INC.

XX Putman RJ, Rao AG, Rao A;

XX WPI: 1995-231570/30.

XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 PT used against plant pathogenic fungi or human or animal infections

XX Claim 1; Page 30; 45pp; English.

XX AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.

XX Sequence 17 AA;

Query Match 65.6%; Score 59; DB 16; Length 17;  
 Best Local Similarity 76.9%; Pred. No. 0.0011;  
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KWAERVAYRGIRY 13  
 || ||| ||||:|  
 Db 1 kwfrvryrgiky 13

## RESULT 10

AAY69609  
 ID AAY69609 standard; peptide; 17 AA.

XX AC AAY69609;

XX

08-MAY-2000 (first entry)

Generic tachyplesin (TP) analogue antimicrobial peptide.

Tachyplesin analogue; generic; antimicrobial; disulphide bond;  
 antifungal; antiviral; antimicrobial; transgenic plant.

XX Synthetic.

XX Key Location/Qualifiers

PH Misc-difference 3

FT /label= Ile, Val, Met, Phe, Tyr

FT Misc-difference 7

FT /label= Ile, Val, Met, Phe, Tyr

FT Misc-difference 12

FT /label= Ile, Val, Met, Phe, Tyr

FT Misc-difference 16

FT /label= Ile, Val, Met, Phe, Tyr

FT /note= "The molecule has the same amino acid at all four  
 of the above positions"

FT Modified-site 17

FT /note= "C-terminal amide"

XX US6015941-A.

XX 18-JAN-2000.

XX 31-OCT-1997; 97US-0962034.

XX 31-OCT-1997; 97US-0962034.

XX (PION-) PIONEER HI-BRED INT INC.

XX Rao AG;

XX WPI: 2000-126327/11.

XX New tachyplesin analogs useful for controlling fungal and bacterial  
 activity in agricultural and medical applications and for controlling  
 plant viruses have four cysteine substitutions -  
 Claim 1; Column 23; 17pp; English.

XX This sequence represents a generic tachyplesin (TP) analogue which has  
 CC antimicrobial activity. Tachyplesin (AAY69608) is a naturally occurring  
 CC antimicrobial peptide which contains two disulphide bonds which help  
 CC to maintain its tertiary structure. The invention relates to novel  
 CC peptide analogues of tachyplesin (Y696912-AAY69614, AAY69616) in which  
 CC the cysteine residues at positions 3, 7, 12 and 16 of the native  
 CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,  
 CC valine, methionine, phenylalanine or tyrosine, the same amino acid being  
 CC present at all four positions. Despite being unable to form  
 CC intramolecular disulphide bonds, the analogues are functional as  
 CC antimicrobial agents. The tachyplesin analogues are useful for  
 CC controlling fungal and viral activity in agricultural and medical  
 CC applications and for controlling plant viruses. They can also be  
 CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,  
 CC soya or especially maize plants to provide resistance to pathogenic fungi  
 CC and viruses.

XX Sequence 17 AA;

Query Match 65.6%; Score 59; DB 21; Length 17;  
 Best Local Similarity 76.9%; Pred. No. 0.0011;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KWAERVAYRGIRY 13

|| ||| ||||:|

Db 1 kwfrvryrgiky 13

RESULT 11

AAR75816  
ID AAR75816 standard; peptide: 21 AA.

XX  
AC AAR75816;

XX  
DT 07-FEB-1996 (first entry)

XX  
DE Antimicrobial tachyplesin peptide derivative.

XX  
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;

XX  
OS infection.

XX  
PN Synthetic.

XX  
PN WO9516776-A1.

XX  
PD 22-JUN-1995.

XX  
PF 19-DEC-1994; 94WO-US14619.

XX  
PR 17-DEC-1993; 93US-0168809.

XX  
PA (PION-) PIONEER HI-BRED INT INC.

XX  
PI Putman RJ, Rao AG, Rao A;

XX  
DR WPI; 1995-231570/30.

XX  
PT New peptide derivs. of tachyplesin - having antimicrobial activity,

XX  
PS used against plant pathogenic fungi or human or animal infections

XX  
PS Claim 1; Page 34; 45pp; English.

XX  
CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
CC haemocytes which has antimicrobial properties. The peptide derivatives  
CC also have antimicrobial activity and can be used for killing and  
CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
CC also be used for treating and preventing infection in humans and  
CC animals.

XX  
SQ Sequence 21 AA;

Query Match 65.6%; Score 59; DB 16; Length 21;

Best Local Similarity 76.9%; Pred. No. 0.0013;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13

DB 1 kwlfrvnyrgiky 13

RESULT 12

AAR75810

ID AAR75810 standard; peptide: 35 AA.

XX  
AC AAR75810;

XX  
DT 07-FEB-1996 (first entry)

XX  
DE Antimicrobial tachyplesin peptide derivative.

XX  
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;

XX  
KW infection.

XX  
OS Synthetic.

XX  
PN WO9516776-A1.

XX  
PD 22-JUN-1995.

XX  
PF 19-DEC-1994; 94WO-US14619.

XX  
PR 17-DEC-1993; 93US-0168809.

XX  
PA (PION-) PIONEER HI-BRED INT INC.

XX  
PI Putman RJ, Rao AG, Rao A;

XX  
DR WPI; 1995-231570/30.

XX  
PT New peptide derivs. of tachyplesin - having antimicrobial activity,

XX  
PS used against plant pathogenic fungi or human or animal infections

XX  
PS Claim 1; Page 31; 45pp; English.

XX  
CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
CC haemocytes which has antimicrobial properties. The peptide derivatives  
CC also have antimicrobial activity and can be used for killing and  
CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
CC also be used for treating and preventing infection in humans and  
CC animals.

XX  
SQ Sequence 35 AA;

Query Match 65.6%; Score 59; DB 16; Length 35;

Best Local Similarity 76.9%; Pred. No. 0.0023;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13

DB 1 kwlfrvnyrgiky 13

RESULT 13

AAR75822

ID AAR75822 standard; peptide: 17 AA.

XX  
AC AAR75822;

XX  
DT 07-FEB-1996 (first entry)

XX  
DE Antimicrobial tachyplesin peptide derivative.

XX  
KW Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;

XX  
KW infection.

XX  
OS Synthetic.

XX  
PN WO9516776-A1.

XX  
PD 22-JUN-1995.

XX  
PF 19-DEC-1994; 94WO-US14619.

XX  
PR 17-DEC-1993; 93US-0168809.

XX  
PA (PION-) PIONEER HI-BRED INT INC.

XX  
PI Putman RJ, Rao AG, Rao A;

XX  
DR WPI; 1995-231570/30.

XX  
PT New peptide derivs. of tachyplesin - having antimicrobial activity,

XX  
PS used against plant pathogenic fungi or human or animal infections

XX  
PS Claim 1; Page 36; 45pp; English.

XX  
CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin

CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC *Fusarium graminearum*, *Fusarium moniliforme*, *Sclerotinia sclerotiorum*,  
 CC *Sclerotinia trifoliorum* and *Aspergillus flavus*. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.

XX Sequence 17 AA;

Query Match 64.4%; Score 58; DB 16; Length 17;

Best Local Similarity 76.9%; Pred. No. 0.0016;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13  
 II III IIII I  
 Db 1 KWRVRYRGLEY 13

RESULT 14

AAW99414  
 ID AAW99414 standard; peptide; 17 AA.

XX AAW99414;

XX AC

XX XX

DT 08-JUN-1999 (first entry)

XX Tachyplesin derivative peptide SM2307.

DE Linear; tachyplesin; peptide antibiotic; beta-sheet; secondary structure;

XX disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;

KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;

KW nucleus; blood-brain barrier.

XX Synthetic.

OS WO9907728-A2.

XX PN

XX PD 18-FEB-1999.

XX PF 06-AUG-1998; 98WO-FR01757.

XX PR 12-AUG-1997; 97FR-0010297.

XX PA (SYNT-) SYNT:EM SA.

XX PI Calas B, Chavanieu A, Grassy G, Kaczorek M;

XX DR WPI; 1999-190034/16.

XX PT Derivatives of antibiotic peptides lacking disulfide bridges - used

XX as carriers to deliver active agents into cells

XX Claim 8; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the tachyplesin family of

XX peptide antibiotics. Tachyplesin antibiotics form part of the peptide

XX antibiotic family which contain a beta-sheet secondary structure linked

XX by disulphide bridges. The new derivatives are linear and lack the

XX disulphide bridge. The novel derivatives are used to deliver active

XX agents to an organism, e.g. therapeutic proteins, antibodies (or their

CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 17 AA;

Query Match 64.4%; Score 58; DB 20; Length 17;

Best Local Similarity 76.9%; Pred. No. 0.0024;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13  
 II III IIII I  
 Db 1 KWRVRYRGLEY 13

RESULT 15

AAW91671  
 ID AAP91671 standard; peptide; 17 AA.

XX AAP91671;

XX AC

XX XX

DT 29-JUN-1990 (first entry)

XX New lipopolysaccharide-binding polypeptide(s).

DE Lipopolysaccharide-binding polypeptide; bacterial infections;

XX lipopolysaccharide (LPS) endotoxins; antibacterial agents;

KW LPS-mediated immune disorders; inflammatory disorders;

KW horseshoe crab haemocytes.

XX Horseshoe crab.

OS Key

XX Location/Qualifiers

XX FT Misc-difference 1

XX FT /label=OTHER

XX FT /note="H-Lys"

XX FT Disulfide-bond 3..16

XX FT Disulfide-bond 7..12

XX FT Misc-difference 17

XX FT /label=OTHER

XX FT /note="Arg-OH or Arg-NH2"

XX PN WO8901492-A.

XX XX

XX PD 23-FEB-1989.

XX PF 19-AUG-1988; 88WO-JP00823.

XX XX

XX PR 21-AUG-1987; 87JP-0206258.

XX PA (SEK ) SEIKAGAKU KOGYO KK.

XX PI Nakamura T, Iwanaga S, Ohno M, Miyazaki K;  
 XX WPI; 1989-068854/09.  
 XX New lipo:polysaccharide- binding polypeptide(s) -  
 XX useful for treating bacterial infections and immune and  
 XX inflammatory disorders.  
 XX Claim 2; Page 27; 39pp; English.

XX The lipopolysaccharide-binding polypeptides may be prepd. by either

XX solid-phase peptide synthesis followed by oxidn. to form the disulphide

XX bridges or, subjecting horseshoe crab haemocytes to hypotonic extraction,

XX extracting the residue with acid, and purifying the extract. The

XX polypeptides have high affinity for lipopolysaccharide (LPS) endotoxins

XX and are useful for removing such toxins from fluids, as antibacterial

XX agents, eg active against *Salmonella* spp. and *S. aureus*, and for

XX treating LPS-mediated immune and inflammatory disorders, eg superior

XX tracheobronchial infections, urinary tract infections, bedsores, burns,

XX colitis, cirrhosis, hepatic insufficiency and post-operative

XX complications.

XX Sequence 17 AA;

Query Match 63.3%; Score 57; DB 10; Length 17;

Best Local Similarity 76.9%; Pred. No. 0.0024;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

us-09-485-571-32.rag

Wed Feb 13 07:52:38 2002

Qy 1 KWAFRVAYRGIRY 13  
  || ||| |||| |  
Db 1 kwcfreyrgicy 13

- Search completed: February 12, 2002, 12:30:34  
Job time: 367 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:25 ; Search time 106.12 seconds  
(without alignments)  
3.605 Million cell updates/sec

Title: US-09-485-571-32

Perfect score: 90

Sequence: 1 KNAFRVAYRGIRYLRL 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	74.4	17	1 US-08-168-809-5	Sequence 5, Appli
2	61	67.8	17	1 US-08-168-809-7	Sequence 7, Appli
3	60	66.7	17	1 US-08-168-809-18	Sequence 18, Appl
4	59	65.6	17	1 US-08-168-809-6	Sequence 6, Appli
5	59	65.6	17	3 US-08-962-034-2	Sequence 2, Appli
6	59	65.6	21	1 US-08-168-809-15	Sequence 15, Appl
7	59	65.6	35	1 US-08-168-809-9	Sequence 9, Appli
8	58	64.4	17	1 US-08-168-809-21	Sequence 21, Appl
9	57	63.3	17	1 US-07-926-965-1	Sequence 1, Appli
10	57	63.3	17	1 US-07-876-883-1	Sequence 1, Appli
11	57	63.3	17	1 US-07-876-883-3	Sequence 3, Appli
12	57	63.3	17	1 US-08-168-809-4	Sequence 4, Appli
13	57	63.3	17	1 US-08-426-550-1	Sequence 1, Appli
14	57	63.3	17	1 US-08-426-550-3	Sequence 3, Appli
15	57	63.3	17	3 US-08-962-034-1	Sequence 1, Appli
16	57	63.3	39	1 US-08-168-809-16	Sequence 16, Appl
17	56	62.2	17	1 US-08-168-809-19	Sequence 19, Appl
18	55	61.1	17	1 US-08-168-809-2	Sequence 2, Appli
19	54	60.0	17	1 US-07-876-883-2	Sequence 2, Appli
20	54	60.0	17	1 US-08-168-809-17	Sequence 17, Appl
21	54	60.0	17	1 US-08-426-550-2	Sequence 2, Appli
22	50	55.6	17	1 US-08-168-809-10	Sequence 10, Appl
23	50	55.6	17	1 US-08-168-809-12	Sequence 12, Appl
24	50	55.6	17	1 US-08-168-809-13	Sequence 13, Appl
25	50	55.6	18	1 US-07-876-883-4	Sequence 4, Appli
26	50	55.6	18	1 US-08-282-030-7	Sequence 7, Appli
27	50	55.6	18	1 US-08-426-550-4	Sequence 4, Appli

28	50	55.6	18	5 PCT-US95-10219-7	Sequence 7, Appli
29	50	55.6	19	1 US-08-282-030-8	Sequence 8, Appli
30	50	55.6	19	5 PCT-US95-10219-8	Sequence 8, Appli
31	49	54.4	17	1 US-08-168-809-11	Sequence 11, Appl
32	49	54.4	17	4 US-09-230-180-36	Sequence 36, Appl
33	47	52.2	17	1 US-08-168-809-14	Sequence 14, Appl
34	47	52.2	18	1 US-08-037-777A-1	Sequence 1, Appli
35	47	52.2	18	1 US-07-876-883-5	Sequence 5, Appli
36	47	52.2	18	1 US-08-379-039C-1	Sequence 1, Appli
37	47	52.2	18	1 US-08-426-550-5	Sequence 5, Appli
38	47	52.2	18	2 US-08-459-400-1	Sequence 1, Appli
39	46	51.1	17	1 US-07-856-026B-21	Sequence 21, Appl
40	45	50.0	17	1 US-07-856-026B-3	Sequence 3, Appli
41	45	50.0	17	1 US-07-856-026B-13	Sequence 13, Appl
42	45	50.0	18	1 US-07-856-026B-4	Sequence 4, Appli
43	45	50.0	18	1 US-07-856-026B-14	Sequence 14, Appl
44	45	50.0	18	1 US-07-856-026B-22	Sequence 22, Appl
45	44	48.9	18	1 US-07-856-026B-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1  
US-08-168-809-5  
; Sequence 5, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; APPLICANT: Rao, Aragula G.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLETSIN HAVING  
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Stree  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,809  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roth, Michael J.  
; REGISTRATION NUMBER: 29,342  
; REFERENCE/DOCKET NUMBER: 0173R US  
; TELEPHONE: 515-245-3595  
; TELEFAX: 515-245-3634  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-168-809-5

Query Match 74.4% Score 67; DB 1; Length 17;  
Best Local Similarity 92.3%; Pred. No. 2 ie-05;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNAFRVAYRGIRY 13

DB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
DB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
DB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
DB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
DB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
DB	1	2	3	4	5	6																																																																																														

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-168-809-6

Query Match 65.6%; Score 59; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0005;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KWAERVYRGIRY 13  
||| ||| ||| |||  
Db 1 KWLFRVYRGIRY 13

## RESULT 5

US-08-962-034-2

; Sequence 2, Application US/08962034

; Patent No. 6015941

; GENERAL INFORMATION:

; APPLICANT: Rao, A. Gururaj

; TITLE OF INVENTION: PEPTIDE DERIVATIVES OF TACHYPLESIN

; TITLE OF INVENTION: HAVING ANTIMICROBIAL ACTIVITY

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell Seltzer IP Group of Alston & Bird

; STREET: 3605 Glenwood Ave.

; CITY: Raleigh

; STATE: NC

; COUNTRY: US

; ZIP: 27622-1107

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/962,034

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Spruill, W. Murray

; REGISTRATION NUMBER: 32,943

; REFERENCE/DOCKET NUMBER: 5718-19

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919 420 2202

; TELEFAX: 919 881 3175

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 17 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-962-034-2

Query Match 65.6%; Score 59; DB 3; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0005;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KWAERVYRGIRY 13  
||| ||| ||| |||  
Db 1 KWLFRVYRGIRY 13

## RESULT 6

US-08-168-809-15

; Sequence 15, Application US/08168809  
; Patent No. 5580852  
; GENERAL INFORMATION:  
; APPLICANT: Putnam, Rebecca J.  
; APPLICANT: Rao, Aragula G.  
; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pioneer Hi-Bred International  
; STREET: 700 Capital Square, 400 Locust Stree  
; CITY: Des Moines  
; STATE: IA  
; COUNTRY: USA  
; ZIP: 50309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/168,809  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Roth, Michael J.  
; REGISTRATION NUMBER: 29,342  
; REFERENCE/DOCKET NUMBER: 0173R US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 515-245-3595  
; TELEFAX: 515-245-3634  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-168-309-15

Query Match 65.6%; Score 59; DB 1; Length 21;  
Best Local Similarity 76.9%; Pred. No. 0.00062;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KWAERVYRGIRY 13  
||| ||| ||| |||  
Db 1 KWLFRVYRGIRY 13

## RESULT 7

US-08-168-809-9

; Sequence 9, Application US/08168809

; Patent No. 5580852

; GENERAL INFORMATION:

; APPLICANT: Putnam, Rebecca J.

; APPLICANT: Rao, Aragula G.

; TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING

; TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pioneer Hi-Bred International

; STREET: 700 Capital Square, 400 Locust Stree

; CITY: Des Moines

; STATE: IA

; COUNTRY: USA

; ZIP: 50309

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/168,809  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Roth, Michael J.  
REGISTRATION NUMBER: 29,342  
REFERENCE/DOCKET NUMBER: 0173R US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 515-245-3595  
TELEFAX: 515-245-3634  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-168-809-9

Query Match 65.6%; Score 59; DB 1; Length 35;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KWAFRVYRGIRY 13  
II III IIII I  
Db 1 KWLFRVYRGIRY 13

RESULT 8  
US-08-168-809-21  
Sequence 21, Application US/08168809  
Patent No. 5580852  
GENERAL INFORMATION:  
APPLICANT: Putnam, Rebecca J.  
APPLICANT: Rao, Aragula G.  
TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
TITLE OF INVENTION: INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pioneer Hi-Bred International  
STREET: 700 Capital Square, 400 Locust Stree  
CITY: Des Moines  
STATE: IA  
COUNTRY: USA  
ZIP: 50309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/168,809  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Roth, Michael J.  
REGISTRATION NUMBER: 29,342  
REFERENCE/DOCKET NUMBER: 0173R US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 515-245-3595  
TELEFAX: 515-245-3634  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-168-809-21

Query Match 64.4%; Score 58; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.00074;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVYRGIRY 13  
II III IIII I  
Db 1 KWLFRVYRGIRY 13

RESULT 9  
US-07-926-965-1  
Sequence 1, Application US/07926965  
Patent No. 5416194  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, TAKANORI; IWANAGA, SADAOKI;  
APPLICANT: OHNO, MOTONORI; MIYAZAKI, KYOSURE  
TITLE OF INVENTION: NOVEL POLYPEPTIDE AND METHOD FOR  
TITLE OF INVENTION: PREPARING THE SAME  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIERMAN & MUSERLIAN  
STREET: 600 THIRD AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/926,965  
FILING DATE: 19920807  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/665,819  
FILING DATE: 07-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/348,487  
FILING DATE: 19-APR-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: CHARLES A. MUSERLIAN  
REGISTRATION NUMBER: 19,683  
REFERENCE/DOCKET NUMBER: TSU-4B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-661-8000  
TELEFAX: 212-661-8002  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY:  
MOLECULE TYPE: POLYPEPTIDE  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: HORSESHOE CRAB  
STRAIN: TACHYPLEUS TRIDENTATUS  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE: HEMOCYTE  
CELL LINE:  
ORGANELLE:  
FEATURE:  
NAME/KEY: LPS-binding polypeptide, or LPB

LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: WHEREIN CYS-3 AND CYS-16 CAN  
FORM A DISULFIDE BOND, AND CYS-7 AND CYS-12 CAN FORM A DISULFIDE BOND; WHEREIN CARBOXYL GROUP OF ARG-17 CAN BE AMIDATED; WHERE  
OTHER INFORMATION: BOND; WHEREIN CARBOXYL GROUP OF ARG-17 CAN BE AMIDATED; WHERE  
OTHER INFORMATION: OR ALL OF THE RESIDUES CAN BE PROTECTED WITH PROTECTIVE GROUP  
US-07-926-965-1

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011; 3; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGYR 13  
II III IIIII I  
Db 1 KWCFRVCYRGICY 13

RESULT 10  
US-07-876-883-1  
Sequence 1, Application US/07876883  
Patent No. 5449752  
GENERAL INFORMATION:  
APPLICANT: Fujii, No. 5449752utaka  
APPLICANT: Yamamoto, Naoki  
APPLICANT: Matsumoto, Akiyoshi  
APPLICANT: Waki, Michinori  
TITLE OF INVENTION: No. 5449752el Polypeptides With Affinity To  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/876,883  
FILING DATE: 19920429  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Miller, Charles E.  
REGISTRATION NUMBER: 24,576  
REFERENCE/DOCKET NUMBER: 7568-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-07-876-883-1

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KWAFRVAYRGYR 13  
II III IIIII I  
Db 1 KWCFRVCYRGICY 13

RESULT 12  
US-08-168-809-4  
Sequence 4, Application US/08168809  
Patent No. 5580852  
GENERAL INFORMATION:  
APPLICANT: Putnam, Rebecca J.  
APPLICANT: Rao, Aragula G.  
TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pioneer Hi-Bred International  
STREET: 700 Capital Square, 400 Locust Street  
CITY: Des Moines  
STATE: IA  
COUNTRY: USA  
ZIP: 50309  
COMPUTER READABLE FORM:

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGYR 13  
II III IIIII I  
Db 1 KWCFRVCYRGICY 13

RESULT 11  
US-07-876-883-3  
Sequence 3, Application US/07876883  
Patent No. 5449752  
GENERAL INFORMATION:  
APPLICANT: Fujii, No. 5449752utaka  
APPLICANT: Yamamoto, Naoki  
APPLICANT: Matsumoto, Akiyoshi  
APPLICANT: Waki, Michinori  
TITLE OF INVENTION: No. 5449752el Polypeptides With Affinity To  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/876,883  
FILING DATE: 19920429  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Miller, Charles E.  
REGISTRATION NUMBER: 24,576  
REFERENCE/DOCKET NUMBER: 7568-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-07-876-883-3

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KWAFRVAYRGYR 13  
II III IIIII I  
Db 1 KWCFRVCYRGICY 13

RESULT 12  
US-08-168-809-4  
Sequence 4, Application US/08168809  
Patent No. 5580852  
GENERAL INFORMATION:  
APPLICANT: Putnam, Rebecca J.  
APPLICANT: Rao, Aragula G.  
TITLE OF INVENTION: DERIVATIVES OF TACHYPLESIN HAVING  
INHIBITORY ACTIVITY TOWARDS PLANT PATHOGENIC FUNGI  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pioneer Hi-Bred International  
STREET: 700 Capital Square, 400 Locust Street  
CITY: Des Moines  
STATE: IA  
COUNTRY: USA  
ZIP: 50309  
COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/168,809  
;; FILING DATE:  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Roth, Michael J.  
;; REGISTRATION NUMBER: 29,342  
;; REFERENCE/DOCKET NUMBER: 0173R US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 515-245-3595  
;; TELEFAX: 515-245-3634  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 17 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; US-08-168-809-4

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13  
|| ||| |||||  
Db 1 KWCRCVYRGICY 13

RESULT 13  
US-08-426-550-1  
; Sequence 1, Application US/08426550  
; Patent No. 5710128  
; GENERAL INFORMATION:  
; APPLICANT: Fujii, No. 5710128utaka  
; APPLICANT: Yamamoto, Naoki  
; APPLICANT: Matsumoto, Akiyoshi  
; APPLICANT: Waki, Michinori  
; TITLE OF INVENTION: Pharmaceutical Compositions of  
; TITLE OF INVENTION: No. 5710128el Lipopolysaccharide-Binding Polypeptides  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,550  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Miller, Charles E.  
; REGISTRATION NUMBER: 24,576  
; REFERENCE/DOCKET NUMBER: 7568-006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-9741  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; US-08-426-550-3

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13  
|| ||| |||||  
Db 1 KWCRCVYRGICY 13

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 17 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: peptide  
;; US-08-426-550-1

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13  
|| ||| |||||  
Db 1 KWCRCVYRGICY 13

RESULT 14  
US-08-426-550-3  
; Sequence 3, Application US/08426550  
; Patent No. 5710128  
; GENERAL INFORMATION:  
; APPLICANT: Fujii, No. 5710128utaka  
; APPLICANT: Yamamoto, Naoki  
; APPLICANT: Matsumoto, Akiyoshi  
; APPLICANT: Waki, Michinori  
; TITLE OF INVENTION: Pharmaceutical Compositions of  
; TITLE OF INVENTION: No. 5710128el Lipopolysaccharide-Binding Polypeptides  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,550  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Miller, Charles E.  
; REGISTRATION NUMBER: 24,576  
; REFERENCE/DOCKET NUMBER: 7568-006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-9741  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; US-08-426-550-3

Query Match 63.3%; Score 57; DB 1; Length 17;  
Best Local Similarity 76.9%; Pred. No. 0.0011;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13  
|| ||| |||||  
Db 1 KWCRCVYRGICY 13

```

RESULT 15
US-08-962-034-1
: Sequence 1, Application US/08962034
: Patent No. 6015941
: GENERAL INFORMATION:
: APPLICANT: Rao, A. Gururaj
: TITLE OF INVENTION: PEPTIDE DERIVATIVES OF TACHYPLESIN
: TITLE OF INVENTION: HAVING ANTIMICROBIAL ACTIVITY
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Bell Seltzer IP Group of Alston & Bird
: STREET: 3605 Glenwood Ave.
: CITY: Raleigh
: STATE: NC
: COUNTRY: US
: ZIP: 27622-1107
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/962,034
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Spruill, W. Murday
: REGISTRATION NUMBER: 32,943
: REFERENCE/DOCKET NUMBER: 5718-19
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919 420 2202
: TELEFAX: 919 881 3175
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 17 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
US-08-962-034-1

```

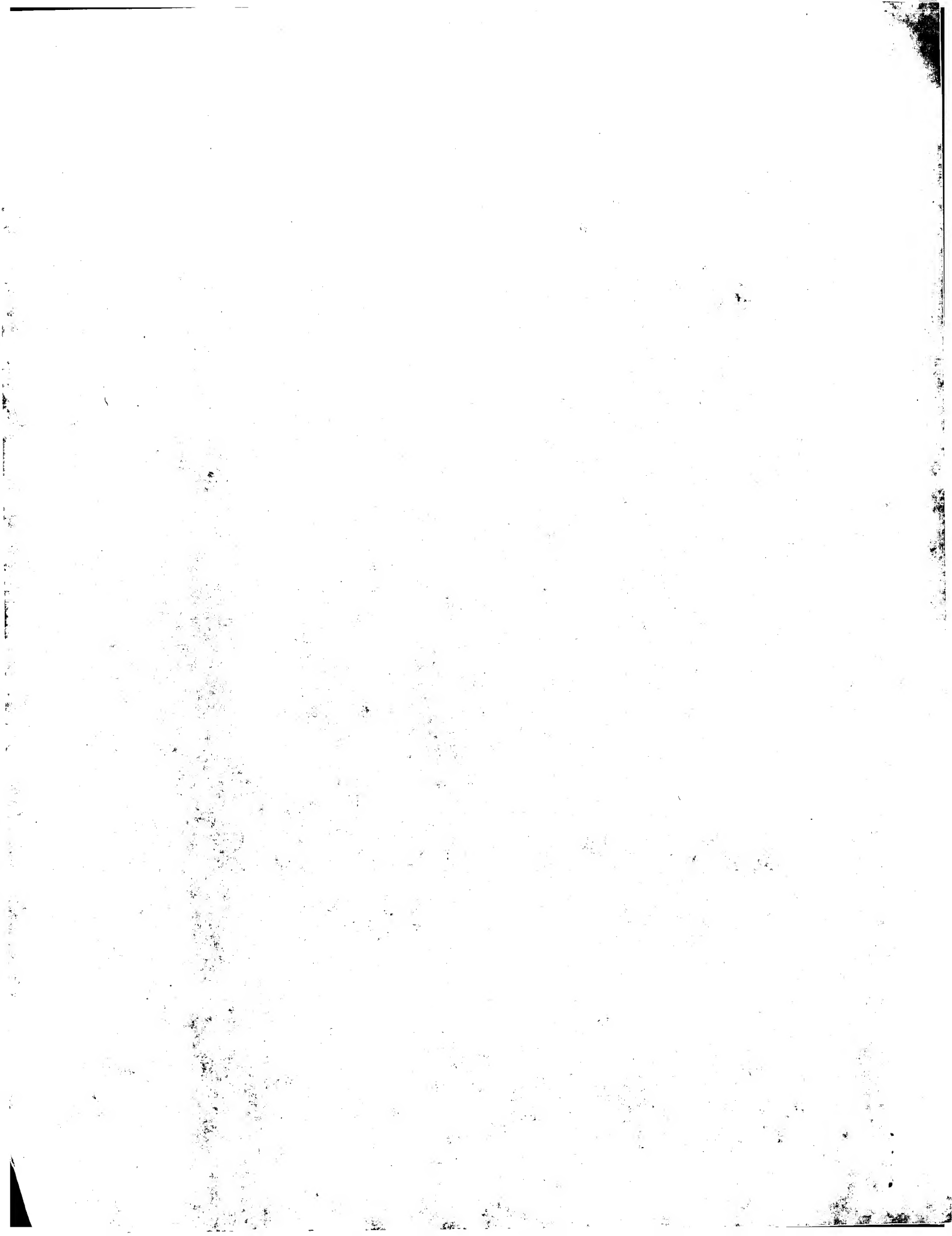
```

Query Match      63.3%; Score 57; DB 3; Length 17;
Best Local Similarity 76.9%; Pred. No. 0.0011;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 KWAERVAYRGIRY 13
Db      1 KWCFRVCYRGICY 13

```

Search completed: February 12, 2002, 12:32:25  
Job time: 453 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:42 ; Search time 126.85 Seconds  
(without alignments)  
10.209 Million cell updates/sec

Title: US-09-485-571-32

Perfect score: 90

Sequence: 1 KWAERVAVRGIRYLRL 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	63.3	17	2 A38824	tachyplesin I - ho
2	57	63.3	17	2 JX0125	tachyplesin III -
3	57	63.3	19	2 JX0124	tachyplesin I prec
4	57	63.3	77	2 A38345	tachyplesin I prec
5	54	60.0	77	2 B38345	tachyplesin II pre
6	50	55.6	18	2 JX0124	polyphemus I - A
7	47	52.2	18	2 JX0125	polyphemus II -
8	44	48.9	307	2 F84162	hypothetical prote
9	44	48.9	378	2 A84161	hypothetical prote
10	44	48.9	1661	2 T21986	hypothetical prote
11	44	48.9	1663	2 T21993	hypothetical prote
12	43	47.8	302	2 S76323	hypothetical prote
13	43	47.8	467	2 T43341	alpha-1,3-mannosyl
14	43	47.8	611	2 G69860	hypothetical prote
15	43	47.8	702	2 T16401	hypothetical prote
16	42	46.7	90	2 T41548	alpha-1,3-mannosyl
17	41	45.6	174	2 T10268	atpase inhibitor p
18	41	45.6	285	2 S77356	granulocyte colony
19	40	44.4	174	2 A26912	hypothetical prote
20	40	44.4	521	2 E64862	gamma-crystallin 5
21	40	44.4	529	2 D71489	probable membrane
22	40	44.4	529	2 H81650	hypothetical prote
23	40	44.4	767	2 G86476	conserved hypothet
24	40	44.4	1165	2 S58236	protein F1504.37 l
25	40	44.4	2626	1 I38344	pyruvate (flavodox
26	39.5	43.9	665	2 E82506	titin, cardiac mus
27	39.5	43.9	665	2 E82506	proteinase II VCA0
28	39	43.3	284	2 S27931	hypothetical prote
29	39	43.3	298	2 T29365	Env/v-mpl fusion p
					hypothetical prote

30	39	43.3	300	2 A69351	heterodisulfide re
31	39	43.3	357	2 T25777	hypothetical prote
32	39	43.3	516	2 E70779	probable glpDl pro
33	39	43.3	553	2 S75892	probable glycerol-
34	39	43.3	625	2 S35317	hematopoietic grow
35	39	43.3	626	2 S37622	proto-oncogene - m
36	39	43.3	1152	2 H86486	protein Tyl/copia-
37	39	43.3	1232	2 B70556	probable respirato
38	38.5	42.8	571	2 T31627	hypothetical prote
39	38	42.2	174	2 B72239	conserved hypothet
40	38	42.2	182	2 F85831	probable transfera
41	38	42.2	182	2 E64971	hypothetical prote
42	38	42.2	293	2 T49172	hypothetical prote
43	38	42.2	358	2 H75264	hypothetical prote
44	38	42.2	465	2 C42603	phospho-beta-gluco
45	38	42.2	514	2 B72752	hypothetical prote

ALIGNMENTS

RESULT 1

A38824

tachyplesin I - horseshoe crab (Tachyplesus gigas)

C:Species: Tachyplesus gigas

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997

C:Accession: A38824

R:Muta T.: Fujimoto, T.; Nakajima, H.; Iwanaga, S.

J. Biochem. 108, 261-266, 1990

A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (

ssing intermediate of its precursor.

A:Reference number: JX0124; MUID:91035357

A:Accession: A38824

A:Molecule type: protein

A:Residues: 1-17 <MUT>

A:Experimental source: hemocyte

C:Keywords: amidated carboxyl end

F:3-16.7-12/Disulfide bonds: #status predicted

F:17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 63.3%; Score 57; DB 2; Length 17;

Best Local Similarity 76.9%; Pred. No. 0.0017;

Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KWAERVAVRGIRY 13

Db 1 KWCPRVCYRGICY 13

RESULT 2

JX0125

tachyplesin III - horseshoe crab (Tachyplesus gigas)

C:Species: Tachyplesus gigas

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997

C:Accession: JX0125

R:Muta T.: Fujimoto, T.; Nakajima, H.; Iwanaga, S.

J. Biochem. 108, 261-266, 1990

A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (

ssing intermediate of its precursor.

A:Reference number: JX0124; MUID:91035357

A:Accession: JX0125

A:Molecule type: protein

A:Residues: 1-17 <MUT>

A:Experimental source: hemocyte

C:Keywords: amidated carboxyl end

F:3-16.7-12/Disulfide bonds: #status predicted

F:17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 63.3%; Score 57; DB 2; Length 17;

Best Local Similarity 76.9%; Pred. No. 0.0017;

Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAVRGIRY 13  
 || ||| |||| I  
 Db 1 KWCFRVCYRGICY 13

RESULT 3  
 JX0124  
 tachyplesin I precursor - horseshoe crab (Carcinoscorpius rotundicauda)  
 C:Species: Carcinoscorpius rotundicauda  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jul-1997  
 C:Accession: JX0124  
 R:Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.  
 J. Biochem. 108, 261-266, 1990  
 A:Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (Carcinoscorpius rotundicauda) and their precursor.  
 A:Reference number: JX0124; MUID:91035357  
 A:Accession: JX0124  
 A:Molecule type: protein  
 A:Residues: 1-19 <MUT>  
 A:Experimental source: hemocyte  
 C:Keywords: amidated carboxyl end  
 F:1-17/Product: tachyplesin I #status experimental <MUT>  
 F:3-16,7-12/Disulfide bonds: #status predicted  
 F:17/Modified site: amidated carboxyl end (Arg) (amide in mature form from following glycosylation)

Query Match 63.3%; Score 57; DB 2; Length 19;  
 Best Local Similarity 76.9%; Pred. No. 0.002;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAVRGIRY 13  
 || ||| |||| I  
 Db 1 KWCFRVCYRGICY 13

RESULT 4  
 A38345  
 tachyplesin I precursor - horseshoe crab (Tachyplesus tridentatus)  
 C:Species: Tachyplesus tridentatus  
 C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 01-Dec-2000  
 C:Accession: A38345; A30068  
 R:Shigenaga, T.; Muta, T.; Toh, Y.; Tokunaga, F.; Iwanaga, S.  
 J. Biol. Chem. 265, 21350-21354, 1990  
 A:Title: Antimicrobial tachyplesin peptide precursor. cDNA cloning and cellular localization of the peptide.  
 A:Reference number: A38345; MUID:91065956  
 A:Accession: A38345  
 A:Molecule type: mRNA  
 A:Residues: 1-77 <SHI>  
 A:Cross-references: GB:M57242; GB:J05689; NID:g161659; PID:g161660  
 R:Nakamura, T.; Furunaka, H.; Miyata, T.; Tokunaga, F.; Muta, T.; Iwanaga, S.; Niwa, M.; J. Biol. Chem. 263, 16709-16713, 1988  
 A:Title: Tachyplesin, a class of antimicrobial peptide from the hemocytes of the horseshoe crab, Tachyplesus tridentatus.  
 A:Reference number: A30068; MUID:89034158  
 A:Accession: A30068  
 A:Molecule type: protein  
 A:Residues: 24-40 <NAK>

Query Match 63.3%; Score 57; DB 2; Length 77;  
 Best Local Similarity 76.9%; Pred. No. 0.0082;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAVRGIRY 13  
 || ||| |||| I  
 Db 24 KWCFRVCYRGICY 36

RESULT 5  
 B38345  
 tachyplesin II precursor - horseshoe crab (Tachyplesus tridentatus)  
 C:Species: Tachyplesus tridentatus

C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 08-Dec-2000  
 C:Accession: B38345; JU0123  
 R:Shigenaga, T.; Muta, T.; Toh, Y.; Tokunaga, F.; Iwanaga, S.  
 J. Biol. Chem. 265, 21350-21354, 1990  
 A:Title: Antimicrobial tachyplesin peptide precursor. cDNA cloning and cellular localization of the peptide.  
 A:Reference number: A38345; MUID:91065956  
 A:Accession: B38345  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-77 <SHI>  
 A:Cross-references: GB:J05689  
 R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J. Biochem. 106, 663-668, 1989  
 A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin I and tachyplesin II.  
 A:Reference number: A91914; MUID:90110066  
 A:Accession: JU0123  
 A:Molecule type: protein  
 A:Residues: 24-40 <MIY>  
 C:Comment: The peptide is one of the antimicrobial peptides found in the Japanese horseshoe crab, Tachyplesus tridentatus.  
 C:Keywords: amidated carboxyl end  
 F:26-39,30-35/Disulfide bonds: #status predicted  
 F:40/Modified site: amidated carboxyl end (Arg) (amide in mature form from following glycosylation)

Query Match 60.0%; Score 54; DB 2; Length 77;  
 Best Local Similarity 69.2%; Pred. No. 0.027;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KWAFRVAVRGIRY 13  
 : || ||| |||| I  
 Db 24 KWCFRVCYRGICY 36

RESULT 6  
 JU0124  
 polyphemus I - Atlantic horseshoe crab  
 C:Species: Limulus polyphemus (Atlantic horseshoe crab)  
 C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jul-1997  
 C:Accession: JU0124  
 R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J. Biochem. 106, 663-668, 1989  
 A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin I and tachyplesin II.  
 A:Reference number: A91914; MUID:90110066  
 A:Accession: JU0124  
 A:Molecule type: protein  
 A:Residues: 1-18 <MIY>  
 C:Comment: The peptide is one of the antimicrobial peptides in the Atlantic horseshoe crab, Limulus polyphemus.  
 C:Keywords: amidated carboxyl end  
 F:4-17,8-13/Disulfide bonds: #status experimental  
 F:18/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 55.6%; Score 50; DB 2; Length 18;  
 Best Local Similarity 61.5%; Pred. No. 0.029;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWAFRVAVRGIRY 13  
 : || ||| |||| I  
 Db 2 KWCFRVCYRGICY 14

RESULT 7  
 JU0125  
 polyphemus II - Atlantic horseshoe crab  
 C:Species: Limulus polyphemus (Atlantic horseshoe crab)  
 C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 11-Jul-1997  
 C:Accession: JU0125  
 R:Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, J. Biochem. 106, 663-668, 1989  
 A:Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin I and tachyplesin II.  
 A:Reference number: A91914; MUID:90110066  
 A:Accession: JU0125  
 A:Molecule type: protein

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:42 ; Search time 126.85 Seconds  
(without alignments)  
10.209 Million cell updates/sec

Title: US-09-485-571-33  
Perfect score: 93  
Sequence: 1 KYAWRVAHRGIRWLLRX 17  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	48.4	2083	2 T42721	CRP-ductin-alpha p
2	44	47.3	481	1 S57294	carboxypeptidase D
3	44	47.3	567	2 F81682	phosphoenolpyruvat
4	43	46.2	125	2 T02519	hypothetical prote
5	43	46.2	183	2 B83183	hypothetical prote
6	43	46.2	221	2 T29509	hypothetical prote
7	42	45.2	311	1 PABY1	phosphoprotein pho
8	42	45.2	313	2 C83874	arsenical pump-dri
9	41.5	44.6	508	1 JC5713	25-hydroxyvitamin
10	41.5	44.6	622	2 D82055	conserved hypothet
11	41	44.1	345	2 T46962	8-oxoguanine DNA-g
12	41	44.1	371	2 S04880	probable transposa
13	41	44.1	372	1 Q0EC47	probable transposa
14	41	44.1	571	2 C71528	probable pps pep p
15	41	44.1	1009	2 S20538	chitin synthase (E
16	40.5	43.5	464	2 F69587	L-arabinose transp
17	40	43.0	147	2 C65134	hypothetical 16.9k
18	40	43.0	193	2 B86016	hypothetical prote
19	40	43.0	197	1 Q0BED6	HHRF6 protein - hu
20	40	43.0	256	2 E75401	hypothetical prote
21	40	43.0	387	1 TVBEI3	42.9k protein kina
22	40	43.0	406	2 T50894	hydroxyneurosporen
23	40	43.0	431	2 E81357	glutamate--trNA li
24	40	43.0	502	2 A86019	probable xylulose
25	40	43.0	536	2 T41652	hypothetical prote
26	40	43.0	1594	2 T30549	hensin - rabbit
27	39	41.9	44	2 S55073	env polyprotein -
28	39	41.9	49	2 S55072	env polyprotein -
29	39	41.9	108	2 T30659	hypothetical prote

thiopurine methylt  
aldo/keto reductas  
ubiquinol--cytochr  
ubiquinol--cytochr  
ubiquinol--cytochr  
ubiquinol--cytochr  
signal-transducin  
beta-glucosidase (  
conserved hypothet  
probable excision  
AMP deaminase (EC  
hypothetical prote  
AMP deaminase (EC  
cell division cont  
env protein gp120(  
env polyprotein pr

ALIGNMENTS

RESULT 1  
T42721  
CRP-ductin-alpha precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 23-Mar-2001  
C:Accession: T42721  
R:Cheng, H.; Bjerknes, M.; Chen, H.  
Anat. Rec. 244, 327-343, 1996  
A:Title: CRP-ductin: a gene expressed in intestinal crypts and in pancreatic and he  
A:Reference number: 222241; MUID:96362470  
A:Accession: T42721  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2083 <CHE>  
A:Cross-references: EMBL:U37438; NID:gl276646; PID:gl276647; PIDN:AAC52505.1  
A:Experimental source: strain BALB/c; jejunal epithelial cells  
C:Keywords: transmembrane protein  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-2083/Product: CRP-ductin-alpha #status predicted <MAT>

Query Match 48.4%; Score 45; DB 2; Length 2083;  
Best Local Similarity 61.5%; Pred. No. 86;  
Matches 8; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 2 YAWRVAHRGIRWL 14  
DB 112 YLMRCSHRG--WL 122

RESULT 2  
S57294  
carboxypeptidase D (EC 3.4.16.6) - Penicillium janthinellum  
N:Alternate names: penicillopeptidase S3  
C:Species: Penicillium janthinellum  
C:Date: 24-Nov-1999 #sequence\_revision 24-Nov-1999 #text\_change 24-Nov-1999  
C:Accession: S57294; S57293  
R:Svendsen, I.; Day, E.S.  
FEBS Lett. 371, 1-3, 1995  
A:Title: The primary structure of carboxypeptidase S3 from Penicillium janthinellum  
A:Reference number: S57294; MUID:95394135  
A:Accession: S57294  
A:Molecule type: protein  
A:Residues: 1-481 <SVE>  
A:Experimental source: strain IBI 3991  
R:Svendsen, I.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S57293  
A:Accession: S57293  
A:Molecule type: protein  
A:Residues: 1-481 <SW>  
A:Experimental source: strain IBI 3991

C:Superfamily: serine carboxypeptidase  
 C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase  
 F:1-481/Product: carboxypeptidase D #status experimental <MAT>  
 F:41,218,256,326,384,392/Binding site: carboxylate (Asn) (covalent) #status experimental  
 F:71-332,191-289,232-279/Disulfide bonds: #status experimental  
 F:158,378,456/Active site: Ser, Asp, His #status predicted

Query Match 47.3%; Score 44; DB 1; Length 481;  
 Best Local Similarity 50.0%; Pred. No. 30;  
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 KYAWRAHGRIRWLLR 16  
 : : ||||| : |||:  
 DB 460 QYQPRVAVRHLEWLLK 475

RESULT 3  
 F81682  
 phosphoenolpyruvate-protein phosphotransferase TC0613 [imported] - Chlamydia muridarum (C.  
 muridarum)  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 21-Jul-2000  
 C:Accession: F81682  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; MUID:20150255  
 A:Accession: F81682  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-567 <TE>  
 A:Cross-references: GB:AE002330; GB:AE002160; NID:g7190649; PIDN:AAF39444.1; PID:g719065  
 A:Experimental source: strain Nigg (MoPn)  
 C:Genetics:  
 A:Gene: TC0613  
 C:Superfamily: phosphotransferase system enzyme I; phosphotransferase system enzyme I hc

Query Match 47.3%; Score 44; DB 2; Length 567;  
 Best Local Similarity 77.8%; Pred. No. 35;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 AHRGIRWLL 15  
 ||| :|||  
 DB 360 AHRSVRWLL 368

RESULT 4  
 T02519  
 hypothetical protein At2g37530 [imported] - Arabidopsis thaliana  
 N:Alternate names: hypothetical protein Fl3M22.3  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 02-Mar-2001  
 C:Accession: T02519; H84793  
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
 submitted to the EMBL Data Library, June 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC Fl3M22 genomic sequence.  
 A:Reference number: Z14677  
 A:Accession: T02519  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-125 <RO>  
 A:Cross-references: EMBL:AC004684; NID:g3236234; PID:g3236236  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487  
 A:Accession: H84793  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-125 <STO>  
 A:Cross-references: GB:AE002093; NID:g3236236; PIDN:AAC23624.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: Fl3M22.3; At2g37530  
 A:Map position: 2  
 A:Introns: 19/2; 67/2  
 C:Superfamily: Arabidopsis thaliana hypothetical protein At2g37530

Query Match 46.2%; Score 43; DB 2; Length 125;  
 Best Local Similarity 47.1%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY 1 KYAWRAH--RGIRWLL 15  
 : : ||| : |||  
 DB 19 KFAWRIPHFVYGATWTL 35

RESULT 5  
 B83183  
 hypothetical protein PA3698 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: B83183  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;  
 Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: B83183  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-183 <STO>  
 A:Cross-references: GB:AE004789; GB:AE004091; NID:g9949857; PIDN:AAG07086.1; GSPDB:G  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA3698

Query Match 46.2%; Score 43; DB 2; Length 183;  
 Best Local Similarity 50.0%; Pred. No. 17;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 YAWRAHGRIRWLL 15  
 : : ||| : |||  
 DB 80 YGFTVSHQGRWYL 93

RESULT 6  
 T29509  
 hypothetical protein W05H7.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T29509  
 R:Wohlmann, P.; Lee, T.T.  
 submitted to the EMBL Data Library, April 1997  
 A:Description: The sequence of C. elegans cosmid W05H7.  
 A:Reference number: Z20630  
 A:Accession: T29509  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-221 <WOH>  
 A:Cross-references: EMBL:U97552; PIDN:AAB52865.1; GSPDB:GN00028; CESP:W05H7.5  
 A:Experimental source: strain Bristol N2; clone W05H7  
 C:Genetics:  
 A:Gene: CESP:W05H7.5  
 A:Map position: X  
 A:Introns: 30/1; 61/1; 172/3; 183/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein W05H7.5

A;Gene: BH1795

Matches 8; Conservative 3; Mismatches 3

A: Experimental source: strain C-125  
C: Genetics:  
A: Gene: BH1795

probable transposase T31 - Escherichia coli insertion sequence IS186  
N:Alternate names: protein b2394  
C:Species: Escherichia coli  
C:Date: 31-Mar-1988 #sequence\_revision 13-Feb-1998 #text\_change 16-Jun-2000  
C:Accession: G65013; D64791; H64721; S40538; A24221  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: G65013  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-372 <BLAT>  
A:Cross-references: GB:AE000327; GB:U00096; NID:g1788731; PIDN:AACT5453.1; PID:g1788  
A:Experimental source: strain K-12, substrain MG1655  
A:Accession: D64791  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 3-372 <BLA2>  
A:Cross-references: GB:AE000163; GB:U00096; NID:g1786790; PIDN:AACT3683.1; PID:g1786  
A:Experimental source: strain K-12, substrain MG1655, insertion sequence IS186  
A:Genetics: Y12  
A:Accession: H64721  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 3-372 <BLA2>  
A:Cross-references: GB:AE000112; GB:U00096; NID:g1786192; PIDN:AACT3127.1; PID:g1786  
A:Experimental source: strain K-12, substrain MG1655, insertion sequence IS186  
A:Genetics: Y11  
R:Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mi  
submitted to the EMBL Data Library, December 1992  
A:Description: Systematic sequencing of the Escherichia coli genome: analysis of the  
A:Reference number: S40531  
A:Accession: S40538  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 3-372 <YUR>  
A:Cross-references: EMBL:D10483; NID:g216434; PIDN:BAA01293.1; PID:g216442  
R:Kochary, R.K.; Jones, D.; Candido, E.P.M.  
J. Bacteriol. 164, 957-959, 1985  
A:Title: IS186: an Escherichia coli insertion element isolated from a cDNA library.  
A:Reference number: A25031; MUID:86033672  
A:Accession: A25031  
A:Molecule type: mRNA  
A:Residues: 3-122 'G', 124-127, 'T', 129-355, 'SHRWIS', 362-363, 'EVIRKEELTRCGE' <KOT>  
A:Cross-references: GB:M11300; NID:g149057; PIDN:AAA25030.1; PID:g149058  
A:Experimental source: insertion sequence IS186  
R:Chong, P.; Hui, I.; Loo, T.; Gillam, S.  
FEBS Lett. 192, 47-52, 1985  
A:Title: Structural analysis of a new GC-specific insertion element IS186.  
A:Reference number: A24221; MUID:86030702  
A:Accession: A24221  
A:Molecule type: mRNA  
A:Residues: 3-355, 'SHRWIS', 362-363, 'EYDPKRTNSLRITKIVMSLOVAIRGTVSLTAYKTOLKNARHL'  
A:Cross-references: GB:X031133; NID:g43689; PIDN:CAA26900.1; PID:g43690  
A:Experimental source: insertion sequence IS186  
C:Comment: This protein is encoded by an insertion sequence isolated from a cDNA lib  
erichia coli.  
C:Genetics: <Y12>  
A:Gene: y181\_2  
A:Mobile element: insertion sequence IS186  
C:Genetics: <Y11>  
A:Gene: y181\_1  
A:Mobile element: insertion sequence IS186  
C:Superfamily: IS186 probable transposase T31  
C:Keywords: transmembrane protein  
F:336-352/Domain: transmembrane #status predicted <TM>

Query Match 44.1%; Score 41; DB 1; Length 372;  
Best Local Similarity 61.5%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Query 2 YAWRVAHGRIRWL 14  
Db 191 YIVRVHWRGLRWL 203

RESULT 13  
QOEC47

probable transposase I (insertion sequence IS421) - Escherichia coli  
N:Alternate names: probable membrane protein I  
C:Species: Escherichia coli  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Sep-1999  
C:Accession: S04880  
R:Sato, S.; Nakada, Y.; Shiratsuchi, A.  
FEBS Lett. 249, 21-26, 1989  
A:Title: IS421, a new insertion sequence in Escherichia coli.  
A:Reference number: S04268; MUID:89252065  
A:Accession: S04880  
A:Molecule type: DNA  
A:Residues: 1-371 <SAT>  
A:Cross-references: EMBL:Y07501; NID:g41836; PIDN:CAA68802.1; PID:g41837  
A:Experimental source: strain K-12  
A:Note: translation of the nucleotide sequence is not complete  
A:Note: translation of the nucleotide sequence from Fig. 3 is inconsistent with the nucl  
C:Genetics:  
A:Mobile element: insertion sequence IS421  
C:Superfamily: IS186 probable transposase T31  
C:Keywords: transmembrane protein  
F:336-352/Domain: transmembrane #status predicted <TM>

Query Match 44.1%; Score 41; DB 2; Length 371;  
Best Local Similarity 61.5%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Query 2 YAWRVAHGRIRWL 14  
Db 191 YIVRVHWRGLRWL 203

RESULT 13  
QOEC47

8-oxoguanine DNA-glycosylase [validated] - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 16-Feb-2001  
C:Accession: T46962  
R:ROSENQUIST, T.A.; ZHARKOV, D.O.; GROLLMAN, A.P.  
Proc. Natl. Acad. Sci. U.S.A. 94, 7429-7434, 1997  
A:Title: Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase.  
A:Reference number: 208708; MUID:97352815  
A:Accession: T46962  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-345 <ROS>  
A:Cross-references: EMBL:U96711; NID:g2078295; PIDN:AAB81133.1; PID:g2078296  
A:Genetics:  
A:Note: mogg1  
C:Function:  
A:Description: removes mutagenic base 8-oxoguanine, formed by oxidative DNA damage, by b  
C:Superfamily: mouse 8-oxoguanine DNA-glycosylase

Query Match 44.1%; Score 41; DB 2; Length 345;  
Best Local Similarity 41.7%; Pred. No. 63;  
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Query 2 YAWRVAHGRIRWL 13  
Db 270 HWQIARHDYCW 281

RESULT 12  
S04880

Search completed: February 12, 2002, 12:34:43  
Job time: 361 sec

```

Qy      2 YANRVHRGIRWL 14
      | | | | | | | |
Db      192 YIVRVHWRGLRWL 204

RESULT 14
C71528
probable pts pep phosphotransferase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Jan-2000
C:Accession: C71528
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A:Reference number: A71570; MUID:99000809
A:Accession: C71528
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <ARN>
A:Cross-references: GB:AE001306; GB:AE001273; NID:G3328748; PIDN:AAC67931.1; PID:G332875
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: ptsI
C:Superfamily: phosphotransferase system enzyme I; phosphotransferase system enzyme I ho
F:17-565/Domain: phosphotransferase system enzyme I homology <ptI>

      Query Match          44.1%; Score 41; DB 2; Length 571;
      Best Local Similarity 66.7%; Pred. No. 1e+02;
      Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      7 AHRGIRWLL 15
      | | | | |
Db      364 SHRSVRWLL 372

RESULT 15
S20538
chitin synthase (EC 2.4.1.16) CHS2 - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Aug-1999
C:Accession: S20538; A38192
R:Chen-Wu, J.L.; Zwicker, J.; Bowen, A.R.; Robbins, P.W.
Mol. Microbiol. 6, 497-502, 1992
A:Title: Expression of chitin synthase genes during yeast and hyphal growth phases of Ca
A:Reference number: S20538; MUID:92219994
A:Accession: S20538
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1009 <CHE>
A:Cross-references: GB:M82937; NID:G170844; PIDN:AAB59308.1; PID:G170845
R:Bowen, A.R.; Chen-Wu, J.L.; Momany, M.; Young, R.; Szaniszló, P.J.; Robbins, P.W.
Proc. Natl. Acad. Sci. U.S.A. 89, 519-523, 1992
A:Title: Classification of fungal chitin synthases.
A:Reference number: A38192; MUID:92115692
A:Accession: A38192
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 323-463, 'FH', 466-510, 'AS', <BOW>
A:Note: sequence extracted from NCBI backbone (NCBIP:75822); this publication is not cit
C:Superfamily: chitin synthase chsA
C:Keywords: glycosyltransferase; hexosyltransferase

      Query Match          44.1%; Score 41; DB 2; Length 1009;
      Best Local Similarity 72.7%; Pred. No. 1.7e+02;
      Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      6 VAHRGIRWLLR 16
      | | | | | | | |
Db      612 VAKRCNWLRL 622

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:54 ; Search time 67.2 Seconds  
(without alignments)  
9.275 Million cell updates/sec

Title: US-09-485-571-33

Perfect score: 93

Sequence: 1 KYAWRVAHGRWLLRX 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	47.3	567	PTL_CHLMU	Q9pk57 chlamydia m
2	42.5	45.7	588	YAV1_SCHPO	Q09758 schizosacch
3	42	45.2	311	PP11_YEAST	P20604 saccharomyc
4	41.5	44.6	508	CP2B_HUMAN	O15528 homo sapien
5	41	44.1	345	OGG1_MOUSE	O07620 mus musculu
6	41	44.1	345	OGG1_RAT	O70249 rattus norv
7	41	44.1	370	INSL_ECOLI	P08409 escherichia
8	41	44.1	371	T421_ECOLI	P11901 escherichia
9	41	44.1	571	PTL_CHLTR	O84340 chlamydia t
10	41	44.1	1009	CHS2_CANAL	P30572 candida alb
11	40	43.0	197	US31_HCMVA	P09707 human cytom
12	40	43.0	387	KR16_HSV1	O00098 ictaluriid h
13	40	43.0	536	YC35_SCHPO	O14061 schizosacch
14	39	41.9	331	GS69_BACSU	P80874 bacillus su
15	39	41.9	388	CYB_MAZE	P04165 zea mays (m
16	39	41.9	392	CYB_SOLTU	P29757 solanum tub
17	39	41.9	392	CYB_VICFA	P05718 vicia faba
18	39	41.9	394	CYB_OENBE	P09843 oenothera b
19	39	41.9	397	CYB_ORYSA	P14833 oryza sativ
20	39	41.9	398	CYB_WHEAT	P07747 triticum ae
21	39	41.9	444	BGLA_THENE	O33843 thermotoga
22	39	41.9	446	BGLA_THEMA	Q08638 thermotoga
23	39	41.9	496	YV31_SCHPO	Q09708 schizosacch
24	39	41.9	836	NOT3_YEAST	P06102 saccharomyc
25	39	41.9	879	AMD2_HUMAN	Q01433 homo sapien
26	39	41.9	880	ENV_SIVNL	P11267 simian immu
27	39	41.9	881	ENV_SIVMK	P05884 simian immu
28	39	41.9	882	ENV_SIVM1	P05885 simian immu
29	39	41.9	896	POL_HTL1C	P14078 human t-cel
30	39	41.9	955	MLL3_MYCLE	O06081 mycobacteri
31	39	41.9	992	IF2B_PYRAB	Q9uzk7 pyrococcus
32	39	41.9	1745	YL51_CAEEL	P34431 caenorhabdi
33	38.5	41.4	460	HEMO_RABIT	P20058 eryctolagus

34	38.5	41.4	653	1	IDUA_HUMAN	P35475 homo sapien
35	38.5	41.4	944	1	CHS2_NEUCR	P30589 neurospora
36	38.5	41.4	1505	1	PK3G_RAT	O70173 rattus norv
37	38	40.9	65	1	LHB2_ECTHL	P11696 ectothiorho
38	38	40.9	246	1	YP73_MYCTU	O50848 mycobacteri
39	38	40.9	267	1	EXOR_RHILV	O52822 rhizobium l
40	38	40.9	268	1	EXOR_RHIME	O52926 rhizobium m
41	38	40.9	352	1	GLN1_DAUCA	O22504 daucus caro
42	38	40.9	367	1	GLNA_CAEEL	P34497 caenorhabdi
43	38	40.9	389	1	OLN1_CHLPN	O07430 chlamydia p
44	38	40.9	400	1	TGT_CAEEL	O23623 caenorhabdi
45	38	40.9	503	1	ATPA_HELPJ	O9zk79 helicobacte

ALIGNMENTS

```
RESULT 1
PTL_CHLMU
ID PTL_CHLMU STANDARD; PRT; 567 AA.
AC O9PK57;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (EC 2.7.3.9)
DE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I).
GN PTSI OR TC0613.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MoPn / Nigg;
RC MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -! FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
CC PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER GROUP (HPR).
CC ENZYME I IS COMMON TO ALL PTS (BY SIMILARITY).
CC -! CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + PROTEIN HISTIDINE -
CC PYRUVATE + PROTEIN N-PHOSPHOHISTIDINE.
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -! SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE002330; AAF39444.1;
CC TIGR; TC0613;
CC InterPro; IPR000121; PEP-utilizers.
CC Pfam; PF000391; PEP-utilizers; 2.
CC ProDom; PD000940; PEP-utilizers; 1.
CC PROSITE; PS00370; PEP-ENZYMES_PHOS_SITE; FALSE_NEG.
CC PROSITE; PS00742; PEP-ENZYMES_2; FALSE_NEG.
CC Phosphotransferase system; Transferase; Kinase; Sugar transport;
CC Phosphorylation; Complete proteome.
CC MOD_RES 205 205
CC SEQUENCE 567 AA; 63311 MW; EFC064A3404F4BB4 CRC64;
CC
```

Query Match	47.3%	Score 44;	DB 1;	Length 567;
Best Local Similarity	77.8%	Pred. No. 11;		
Matches	7;	Conservative	1;	Mismatches
Indels	0;	Gaps	0;	
QY	7	AHRGIRWLL 15		
Db	360	AHRSVRWLL 368		
RESULT	2			
YA7L_SCHPO		STANDARD;	PRT;	588 AA.
ID	YA7L_SCHPO			
AC	Q09758; Q9C025;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	HYPOTHETICAL 69.1 KDA PROTEIN C24H6.01C IN CHROMOSOME I.			
GN	SPAC24H6.01C OR SPAPB21F2.01.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE OF 1-228 FROM N.A.			
RC	STRAIN=972;			
RA	Skellon J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.,			
RA	Wood V.,			
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 211-588 FROM N.A.			
RC	STRAIN=972;			
RA	Saunders D., Harris D., Wood V., Rajandream M.A., Barrell B.G.;			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- SIMILARITY: TO YEAST YGL084C.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL; Z54142; CAA90845.2; -			
DR	EMBL; AL590562; CAC36890.1; -			
KW	Hypothetical protein; Transmembrane.			
FT	TRANSMEM 53 73 POTENTIAL.			
FT	TRANSMEM 120 140 POTENTIAL.			
FT	TRANSMEM 160 180 POTENTIAL.			
FT	TRANSMEM 192 212 POTENTIAL.			
FT	TRANSMEM 301 321 POTENTIAL.			
FT	TRANSMEM 344 364 POTENTIAL.			
FT	TRANSMEM 374 394 POTENTIAL.			
FT	TRANSMEM 445 465 POTENTIAL.			
FT	TRANSMEM 466 486 POTENTIAL.			
FT	TRANSMEM 513 533 POTENTIAL.			
FT	DOMAIN 22 40 SER-RICH.			
FT	SEQUENCE 588 AA; 69089 MW; 757AB685494B8B7A CRC64;			
SQ				
Query Match	45.7%	Score 42.5;	DB 1;	Length 588;
Best Local Similarity	57.1%	Pred. No. 19;		
Matches	8;	Conservative	1;	Mismatches
Indels	4;	Gaps	1;	
QY	4	WRVAHRGI-RWLLR 16		
Db	428	WRWHRSPNRWLR 441		
RESULT	3			
PP1L_YEAST				

ID	PP1L_YEAST	STANDARD;	PRT;	311 AA.
AC	P20604;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	SERINE/THREONINE PROTEIN PHOSPHATASE PPI-1 (EC 3.1.3.16).			
GN	PPH1 OR SIT4 OR YDL047W OR D2693.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE=89136000; PubMed=2537149;			
RA	Arndt K.T., Styles C.A., Fink G.R.;			
RT	"A suppressor of a HIS4 transcriptional defect encodes a protein with			
RL	homology to the catalytic subunit of protein phosphatases.";			
CC	Cell 56:527-537(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C;			
RA	Paulin L., Saren A.M., Laamanen P.;			
RL	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	CHARACTERIZATION.			
RC	MEDLINE=91172202; PubMed=1848673;			
RA	Sutton A., Immanuel D., Arndt K.T.;			
RT	"The SIT4 protein phosphatase functions in late G1 for progression			
RL	into S phase.";			
CC	Mol. Cell. Biol. 11:2133-2148(1991).			
CC	-1- FUNCTION: INVOLVED IN THE DEPHOSPHORYLATION OF THE LARGE SUBUNIT			
CC	OF RNA POLYMERASE II. IS REQUIRED IN LATE G1 FOR NORMAL G1 CYCLIN			
CC	EXPRESSION. BUD INITIATION AND EXPRESSION OF CERTAIN GENES THAT			
CC	ARE PERIODICALLY EXPRESSED DURING LATE G1. ASSOCIATES WITH THE			
CC	SAP PROTEINS IN A CELL CYCLE-DEPENDENT MANNER.			
CC	-1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O -> A PROTEIN +			
CC	ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).			
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC.			
CC	-1- DEVELOPMENTAL STAGE: FUNCTIONS IN THE LATE CELL CYCLE G1 PHASE FOR			
CC	PROGRESSION INTO THE S PHASE, POSSIBLY ASSOCIATED IN TWO SEPARATE			
CC	COMPLEXES WITH THE PHOSPHORYLATED FORMS OF P155 AND P190, TWO HIGH			
CC	MW PROTEINS.			
CC	-1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-V			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL; M24395; AAA56864.1; -			
DR	EMBL; Z71781; CAA96442.1; -			
DR	EMBL; Z74095; CAA98609.1; -			
DR	PIR; A31874; PABY1.			
DR	HSSP; P08129; 1FJM.			
DR	SGD; S0002205; SIT4.			
DR	InterPro; IPR000934; Ser_thr_phosphatse.			
DR	Pfam; PF00149; Stphosphatase; 1.			
DR	PRINTS; PR00114; STPHPTASE.			
DR	SMART; SM00156; PP2AC; 1.			
DR	PROSITE; PS00125; SER_THR_PHOSPHATASE; 1.			
KW	Hydrolase; Iron; Manganese; Cell cycle; Mitosis; Multigene family.			
FT	METAL 53 53 IRON (BY SIMILARITY).			
FT	METAL 55 55 IRON (BY SIMILARITY).			
FT	METAL 82 82 IRON AND MANGANESE (BY SIMILARITY).			
FT	METAL 114 114 MANGANESE (BY SIMILARITY).			
FT	ACT_SITE 115 115 GENERAL ACID (BY SIMILARITY).			
FT	METAL 164 164 MANGANESE (BY SIMILARITY).			
FT	METAL 238 238 MANGANESE (BY SIMILARITY).			
SQ	SEQUENCE 311 AA; 35537 MW; AF52BC65E4E715EC CRC64;			

```
Query Match      45.2%; Score 42; DB 1; Length 311;
Best Local Similarity 58.3%; Pred. No. 12;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 ANRVHRGIRWL 14
DB 205 AQVSPRGAGWL 216

RESULT 4
CP2B_HUMAN STANDARD; PRT; 508 AA.
AC Q15528;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 25-HYDROXYVITAMIN D-1 ALPHA-HYDROXYLASE, MITOCHONDRIAL PRECURSOR
DE (EC 1.14.-.-) (25-OH-D-1 ALPHA-HYDROXYLASE) (P450C1 ALPHA) (P450VD1-
DE ALPHA-HYDROXYLASE) (VD3 1A HYDROXYLASE)
GN CYP27B1 OR CYP27B OR CYP1A1PHA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98089075; PubMed=9428799;
RA Fu G.K., Portale A.P., Miller W.L.;
RT "Complete structure of the human gene for the vitamin D 1alpha-
RT hydroxylase, P450c1alpha."
RL DNA Cell Biol. 16:1499-1507(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=98008873; PubMed=9344864;
RA Monkawa T., Yoshida T., Wakino S., Shinki T., Anazawa H., Deluca H.F.,
RA Suda T., Hayashi M., Saruta T.;
RT "Molecular cloning of cDNA and genomic DNA for human
RT 25-hydroxyvitamin D3 1 alpha-hydroxylase."
RL Biochem. Biophys. Res. Commun. 239:527-533(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98075882; PubMed=9415400;
RA Fu G.K., Lin D., Zhang Y.H., Bikle D.D., Shackleton C.H., Miller W.L.,
RA Portale A.A.;
RT "Cloning of human 25-hydroxyvitamin D-1 alpha-hydroxylase and
RT mutations causing vitamin D-dependent rickets type 1."
RL Mol. Endocrinol. 11:1961-1970(1997).
CC -1- FUNCTION: CATALYZES THE CONVERSION OF 25-HYDROXYVITAMIN D3
CC (25(OH)D) TO 1-ALPHA,25-DIHYDROXYVITAMIN D3 (1,25(OH)2D) PLAYS AN
CC IMPORTANT ROLE IN CRUCIAL ROLE IN NORMAL BONE GROWTH, CALCIUM
CC METABOLISM, AND TISSUE DIFFERENTIATION.
CC -1- PATHWAY: SECOND STEP IN THE CONVERSION OF VITAMIN D(3) INTO THE
CC ACTIVE FORM (1-ALPHA,25-DIHYDROXYVITAMIN D(3)).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- TISSUE SPECIFICITY: KIDNEY.
CC -1- DISEASE: DEFECTS IN CYP27B ARE A CAUSE OF VITAMIN D-DEPENDENT
CC RICKETS TYPE 1 (VDDR-1), A DISEASE CHARACTERIZED BY MUSCLE
CC WEAKNESS AND RICKETS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AF027152; AAC51854.1; -.
```

```
DR EMBL; AB005038; BAA23416.1; -.
DR EMBL; AB005989; BAA22656.1; -.
DR EMBL; AB005990; BAA22657.1; -.
DR EMBL; AB006987; BAA23418.1; -.
DR EMBL; AF020192; AAC51853.1; -.
DR HSSP; P00189; 1SCC.
DR DR MIM; 264700; -.
DR DR InterPro; IPR001128; Cyt_P450.
DR DR Pfam; PF00067; P450; 1.
DR DR PRINTS; PR00385; P450.
DR DR PRINTS; PR00408; MITP450.
DR DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Electron transport; Oxidoreductase; Monooxygenase; Membrane;
KW Heme; Mitochondrion; Transit peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 508 25-HYDROXYVITAMIN D-1 ALPHA HYDROXYLASE.
FT BINDING 455 455 HEME (BY SIMILARITY).
SQ SEQUENCE 508 AA; 56504 MW; 7F0611EFAD1B5C1C CRC64;

Query Match      44.6%; Score 41.5; DB 1; Length 508;
Best Local Similarity 69.2%; Pred. No. 24;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KYARVVAHRGIRW 13
DB 6 KYASRVFHR-VRW 17

RESULT 5
OGGI_MOUSE STANDARD; PRT; 345 AA.
AC O0876C; O08991; O35915; O35617; O08733; O08910; Q9QXE8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE N-GLYCOSYLASE/DNA LYASE [INCLUDES: 8-OXOGUANINE DNA GLYCOSYLASE
DE (EC 3.2.2.-); DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE
DE (EC 4.2.99.18) (AP-LYASE)].
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97352815; PubMed=9207108;
RA Rosenquist T.A., Zharkov D.O., Grollman A.P.;
RT "Cloning and characterization of a mammalian 8-oxoguanine DNA
RT glycosylase."
RL Proc. Natl. Acad. Sci. U.S.A. 94:7429-7434(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97342862; PubMed=9197244;
RA Lu R., Nash H.M., Verdine G.L.;
RT "A mammalian DNA repair enzyme that excises oxidatively damaged
RT guanines maps to a locus frequently lost in lung cancer."
RL Curr. Biol. 7:397-407(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97330655; PubMed=9187114;
RA Aburatani H., Hippi Y., Ishida T., Takashima R., Matsuba C.,
RA Kodama T., Takao M., Yasui A., Yamamoto K., Asano M., Fukasawa K.,
RA Yoshinari T., Inoue H., Otsuka E., Nishimura S.;
RT "Cloning and characterization of mammalian 8-hydroxyguanine-specific
RT DNA glycosylase/apurinic, apyrimidinic lyase, a functional mut
RT homologue."
RL Cancer Res. 57:2151-2156(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HEN;
RX MEDLINE=98096582; PubMed=9434942;
RA Tani M., Shimura K., Kohno T., Takenoshita S., Nagamachi Y.,
RA Yokota J.;
```

RT "Genomic structure and chromosomal localization of the mouse Ogg1 gene  
 RT that is involved in the repair of 8-hydroxyguanine in DNA damage."  
 RL Mamm. Genome 9:32-37(1998).

[5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=98026907; Pubmed=9321410;  
 RA Bjoras M., Luna L., Johnsen B.E., Hoff E., Haug T., Rognes T.,  
 RA Seeberg E.;  
 RT "Opposite base-dependent reactions of a human base excision repair  
 RT enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites."  
 RL EMBO J. 16:6314-6322(1997).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RA Radicella J.P., Rellie F., Dherin C., Boiteux S.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RA Johnsen B., Luna L., Rognes T., Seeberg E.;  
 RT "complete genomic DNA sequence of the Mus musculus 8-oxoguanine DNA  
 RT glycosylase 1 gene (OGH1).";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: DNA REPAIR ENZYME THAT INCISES DNA AT 8-OXOG RESIDUES.  
 CC EXCISES 7,8-DIHYDRO-8-OXOGUANINE AND 2,6-DIAMINO-4-HYDROXY-5-N-  
 CC METHYLFORMADOPYRIMIDINE (FAPY) FROM DAMAGED DNA. HAS A BETA-  
 CC LYASE ACTIVITY THAT NICKS DNA 3' TO THE LESION.  
 CC -!- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE NEAR APURINIC OR  
 CC APYRIMIDINIC SITES TO PRODUCTS WITH 5'-PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS.  
 CC -!- SIMILARITY: BELONGS TO THE OGG1 FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AF003596; AAB61289.1; -;  
 CC EMBL; U88621; AAB68616.1; -;  
 CC EMBL; AF000669; AAB63151.1; -;  
 CC EMBL; AF012916; AAB94512.1; -;  
 CC EMBL; AF012912; AAB94512.1; JOINED.  
 CC EMBL; AF012913; AAB94512.1; JOINED.  
 CC EMBL; AF012914; AAB94512.1; JOINED.  
 CC EMBL; AF012915; AAB94512.1; JOINED.  
 CC EMBL; U96711; AAB81133.1; -;  
 CC EMBL; Y11247; CAA72117.1; -;  
 CC EMBL; Y13479; CAA73883.1; -;  
 CC EMBL; MGJ.1097693; Oqg1.  
 CC MGD; AJ001307; CAB65240.1; -;  
 CC InterPro: IPR003265; Endo\_3c.  
 CC Pfam; PF00730; HhH-GPD; 1.  
 CC SMART; SM00478; ENDO3c; 1.  
 CC Hydroxylase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;  
 CC Multifunctional enzyme; Nuclear protein.  
 CC ACT\_SITE 249 249 BY SIMILARITY.  
 CC CONFLICT 10 10 S -> Q (IN REF. 2).  
 CC CONFLICT 23 25 WAS -> SVA (IN REF. 2).  
 CC CONFLICT 239 239 A -> G (IN REF. 2).  
 CC CONFLICT 299 299 L -> S (IN REF. 6).  
 CC CONFLICT 329 329 S -> F (IN REF. 1).  
 CC CONFLICT 336 336 R -> H (IN REF. 5 AND 7).  
 CC SEQUENCE 345 AA; 38883 MW; 75BB0DD8084E4947 CRC64;

Query Match 44.1%; Score 41; DB 1; Length 345;  
 Best Local Similarity 41.7%; Pred. No. 20;  
 Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 YAWRVAHRGIRW 13  
 : |::||| |  
 DB 270 HWQIAHRDYGW 281

RESULT 6  
 OGG1\_RAT  
 ID OGG1\_RAT STANDARD; PRT; 345 AA.  
 AC O70249;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE N-GLYCOSYLASE/DNA LYASE [INCLUDES: 8-OXOGUANINE DNA GLYCOSYLASE  
 DE (EC 3.2.2.-); DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE  
 DE (EC 4.2.99.18) (AP LYASE)].  
 GN OGG1 OR MMH OR OGH1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hepatoma;  
 RA Prieto-Alamo M.J., Laval F.;  
 RT "Cloning and characterization of a rat 8-oxoguanine-DNA-glycosylase.";  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases  
 CC -!- FUNCTION: DNA REPAIR ENZYME THAT INCISES DNA AT 8-OXOG RESIDUES.  
 CC EXCISES 7,8-DIHYDRO-8-OXOGUANINE AND 2,6-DIAMINO-4-HYDROXY-5-N-  
 CC METHYLFORMADOPYRIMIDINE (FAPY) FROM DAMAGED DNA. HAS A BETA-  
 CC LYASE ACTIVITY THAT NICKS DNA 3' TO THE LESION.  
 CC -!- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE NEAR APURINIC OR  
 CC APYRIMIDINIC SITES TO PRODUCTS WITH 5'-PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE OGG1 FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AF029690; AAC77525.1; -;  
 CC InterPro: IPR003265; Endo\_3c.  
 CC Pfam; PF00730; HhH-GPD; 1.  
 CC SMART; SM00478; ENDO3c; 1.  
 CC Hydroxylase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;  
 CC Multifunctional enzyme; Nuclear protein.  
 CC ACT\_SITE 249 249 BY SIMILARITY.  
 CC SEQUENCE 345 AA; 38711 MW; B7FDF8C782644C41 CRC64;

Query Match 44.1%; Score 41; DB 1; Length 345;  
 Best Local Similarity 41.7%; Pred. No. 20;  
 Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 YAWRVAHRGIRW 13  
 : |::||| |  
 DB 270 HWQIAHRDYGW 281

RESULT 7  
 INSL\_ECOLI  
 ID INSL\_ECOLI STANDARD; PRT; 370 AA.  
 AC P08409; P11307; Q47051; P77426; P76952;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PUTATIVE TRANSPOSASE INSL FOR INSERTION SEQUENCE ELEMENT IS186A/B/C.  
 GN (INSL1 OR B0016) AND (INSL2 OR B0582) AND (INSL3 OR B2394).  
 OS Escherichia coli.



Query Match	44.1%;	Score 41;	DB 1;	Length 571;
Best Local Similarity	66.7%;	Pred. No. 32;		



GS69\_BACSU STANDARD; PRT; 331 AA.  
ID P80874; 007593;  
AC 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GENERAL STRESS PROTEIN 69. (GSP69).  
GN YHON.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Noback M.A., Terpestra P., Holsappel S., Venema G., Bron S.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-25.  
RC STRAIN=IS58;  
RX MEDLINE=97443988; PubMed=9298659;  
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,  
RA Hecker M.;  
RT "First steps from a two-dimensional protein index towards a response-  
RT regulation map for Bacillus subtilis.";  
RL Electrophoresis 18:1451-1463(1997).  
CC -1- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE  
CC LIMITATION AND OXYGEN LIMITATION.  
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y14082; CAA74498.1; -;  
DR EMBL; Z99109; CAB12792.1; -;  
DR HSSP; P23457; 1LWI.  
DR Subtilist; BG13020; yhdN.  
DR Pfam; PF00248; aldo\_ket\_red; 1.  
KW Oxidoreductase; Heat shock; Complete proteome.  
FT ACT\_SITE 125 125 HYDROGEN-BOND DONOR (PROBABLE).  
FT CONFLICT 25 25 G > X (IN REF. 2).  
SQ SEQUENCE 331 AA; 37312 MW; 82BC24D46E4994D0 CRC64;  
  
Query Match 41.9%; Score 39; DB 1; Length 331;  
Best Local Similarity 40.0%; Pred. No. 39;  
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 KYARVAHGRIRWLL 15  
:| | | :|||  
DB 255 RYKSVIHLAVRWIL 269  
  
RESULT 15  
CYB\_MAIZE STANDARD; PRT; 388 AA.  
ID CYB\_MAIZE  
AC P04165;  
DT 20-MAR-1987 (Rel. 04, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CYTOCHROME B.  
GN COB OR CYTB.  
OS Zea mays (Maize).  
OG Mitochondrion.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
  
[1]  
RN SEQUENCE FROM N.A.  
RP Dawson A.J., Jones V.P., Leaver C.J.;  
RT "The apocytochrome b gene in maize mitochondria does not contain  
RT introns and is preceded by a potential ribosome binding site.";  
RL EMBO J. 3:2107-2113(1984).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
CC COUPLED TO ATP SYNTHESIS.  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
CC BOUND TO THE PROTEIN.  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X00789; CAA25367.1; -;  
DR PIR; A00156; CBZM.  
DR MaizeDB; 69227; -;  
DR Mendel; 2200; ZEAna; cob;1.  
DR InterPro; IPR000179; Cyt\_b\_b6.  
DR Pfam; PF00032; cytochrome\_b\_c; 1.  
DR Pfam; PF00033; cytochrome\_b\_n; 1.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00193; CYTOCHROME\_B\_QO; 1.  
DR Electron transport; Mitochondrion; Respiratory chain; Transmembrane;  
KW Heme.  
KW METAL 88 88 IRON 1 (HEME B562 AXIAL LIGAND).  
FT METAL 102 102 IRON 2 (HEME B566 AXIAL LIGAND).  
FT METAL 189 189 IRON 2 (HEME B566 AXIAL LIGAND).  
FT METAL 203 203 IRON 1 (HEME B566 AXIAL LIGAND).  
SQ SEQUENCE 388 AA; 43567 MW; 56CF734C2B62B3F6 CRC64;  
  
Query Match 41.9%; Score 39; DB 1; Length 388;  
Best Local Similarity 53.8%; Pred. No. 46;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 AWRVAHGRIRWLL 15  
:| | | :|||  
DB 323 SFRPIHQGFWLL 335  
  
Search completed: February 12, 2002, 12:39:55  
Job time: 808 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:44 ; Search time 232.64 Seconds  
(without alignments)  
10.689 Million cell updates/sec

Title: US-09-485-571-33  
Perfect score: 93  
Sequence: 1 KYARVAHGRINWLLRX 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	48.4	1957	11 Q9JMJ9	Q9JMJ9 mus musculus
2	45	48.4	2083	11 Q60997	Q60997 mus musculus
3	44	47.3	125	4 Q9HA69	Q9HA69 homo sapien
4	44	47.3	133	2 Q9AF21	Q9AF21 shigella fl
5	44	47.3	298	11 Q9Z0F5	Q9Z0F5 mus musculus
6	44	47.3	481	3 Q9URG3	Q9URG3 penicillium
7	43	46.2	125	10 Q80923	Q80923 arabidopsis
8	43	46.2	183	2 Q9HXU2	Q9HXU2 pseudomonas
9	43	46.2	213	5 Q9VYP6	Q9VYP6 drosophila
10	43	46.2	221	5 Q02175	Q02175 caenorhabdi
11	42.5	45.7	378	3 Q9C0Z5	Q9C0Z5 schizosacch
12	42	45.2	133	2 Q9AFQ8	Q9AFQ8 shigella fl
13	42	45.2	293	5 Q9V3U3	Q9V3U3 drosophila
14	42	45.2	309	3 Q9P4W5	Q9P4W5 kluyveromyc
15	42	45.2	313	2 Q9KX9	Q9KX9 bacillus ha
16	41.5	44.6	65	4 Q9UP42	Q9UP42 homo sapien
17	41.5	44.6	622	2 Q9KNY0	Q9KNY0 vibrio chol
18	41.5	44.6	861	11 Q9EQK5	Q9EQK5 mus musculus
19	41	44.1	144	2 Q69575	Q69575 mycobacteri

20	41	44.1	204	2 Q9L3G5	Q9L3G5 klebsiella
21	41	44.1	208	2 Q9X2V8	Q9X2V8 escherichia
22	41	44.1	435	2 Q9LCS0	Q9LCS0 arthrobacte
23	41	44.1	488	11 Q35405	Q35405 mus musculu
24	41	44.1	711	5 Q9TXJ2	Q9TXJ2 leishmania
25	41	44.1	833	5 Q9N6E9	Q9N6E9 leishmania
26	40.5	43.5	162	5 Q9V849	Q9V849 drosophila
27	40.5	43.5	178	2 Q9LLY0	Q9LLY0 streptomyce
28	40.5	43.5	464	2 P96710	P96710 bacillus su
29	40	43.0	256	2 Q9RUK4	Q9RUK4 deinococcus
30	40	43.0	261	5 Q9NC93	Q9NC93 trypanosoma
31	40	43.0	272	4 Q9S992	Q9S992 homo sapien
32	40	43.0	288	2 Q9KYH2	Q9KYH2 streptomyce
33	40	43.0	304	2 Q9A559	Q9A559 caulobacter
34	40	43.0	310	11 Q9D3W8	Q9D3W8 mus musculu
35	40	43.0	316	4 Q9NUJ5	Q9NUJ5 homo sapien
36	40	43.0	361	10 Q9SNV0	Q9SNV0 antirrhinum
37	40	43.0	406	2 P95619	P95619 rhodocyclu
38	40	43.0	406	2 Q9JPB4	Q9JPB4 rhodocyclu
39	40	43.0	431	2 Q9PP78	Q9PP78 campylobact
40	40	43.0	450	2 P72272	P72272 rhizobium f
41	40	43.0	480	2 Q53032	Q53032 rhodospiril
42	40	43.0	545	12 Q66434	Q66434 dengue viru
43	40	43.0	555	3 P79078	P79078 cryptococcu
44	40	43.0	629	2 Q9F3Q6	Q9F3Q6 streptomyce
45	40	43.0	729	4 Q9UHD2	Q9UHD2 homo sapien

#### ALIGNMENTS

RESULT 1  
Q9JMJ9 PRELIMINARY; PRT; 1957 AA.  
AC Q9JMJ9; 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE VOMERGLANDIN PRECURSOR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DDY;  
RA MEDLINE=20145422; PubMed=10679193;  
RA Matsushita F., Miyawaki A., Mikoshiba K.;  
RT "Vomeroglandin/CRP-ductin is strongly expressed in the glands  
RT associated with the mouse vomeronasal organ: Identification and  
RT characterization of mouse vomeroglandin.";  
RL Biochem. Biophys. Res. Commun. 268:275-281(2000).  
DR EMBL; AB005909; BAA92266.1; -;  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR001190; SRCR.  
DR InterPro; IPR001507; zona\_pellucida.  
DR Pfam; PF00431; CUB; 5.  
DR Pfam; PF00530; SRCR; 7.  
DR Pfam; PF00100; zona\_pellucida; 1.  
DR PRINTS; PR00258; SPERACTRCPTR.  
DR SMART; SM00042; CUB; 5.  
DR SMART; SM00202; SR; 7.  
DR SMART; SM00241; 2P; 1.  
DR PROSITE; PS01180; CUB; 5.  
DR PROSITE; PS00420; SRCR\_1; 7.  
DR PROSITE; PS0287; SRCR\_2; 7.  
DR PROSITE; PS0682; ZP\_DOMAIN; UNKNOWN\_1.  
KW Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 1957  
SQ SEQUENCE 1957 AA; 213178 MW; B3F63F6439EF0D13 CRC64;

Query Match 48.4%; Score 45; DB 11; Length 1957;  
Best Local Similarity 61.5%; Pred. No. 1.7e+02;  
Matches 8; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 2 YAWRVHARGIRWL 14  
DB 123 YLWRCSHRG--WL 133

RESULT 2  
Q60997 PRELIMINARY; PRT; 2083 AA.  
AC Q60997;  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUN-1997 (TReMBLrel. 04, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE CRP-DUCTIN PRECURSOR (CRP).  
GN CRPD OR CRP OR CRP-DUCTIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=JEJUNUM;  
RX MEDLINE=96362470; PubMed=8742698;  
RA Cheng H., Bjerknes M., Chen H.;  
RT "CRP-ductin: a gene expressed in intestinal crypts and in pancreatic  
and hepatic ducts";  
RL Anat. Rec. 244:327-343(1996).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN (CRP-ALPHA AND CRP-  
BETA) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE. THE  
SEQUENCE SHOWN HERE IS THAT OF CRP-ALPHA.  
CC -1- TISSUE SPECIFICITY: INTESTINE AND PANCREAS. IN THE COLON,  
EXPRESSED IN CRYPT CELLS BUT NOT MATURE VILLOS CELLS. MOST HIGHLY  
EXPRESSED IN MID-CRYPT. ALSO EXPRESSED IN EPITHELIUM LINING  
HEPATIC AND PANCREATIC DUCTS.  
CC EMBL; U37438; AAC52505.1; -.  
DR HSP; P29392; ISFP.  
DR MGD; MG1.106210; Crpd.  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR001190; SRCR.  
DR InterPro; IPR001507; zona\_pellucida.  
DR Pfam; PF00431; CUB; 5.  
DR Pfam; PF00530; SRCR; 8.  
DR Pfam; PF00100; zona\_pellucida; 1.  
DR PRINTS; PR00258; SPERACTRCPTR.  
DR SMART; SM00042; CUB; 5.  
DR SMART; SM00202; SR; 8.  
DR SMART; SM00241; ZP; 1.  
DR PROSITE; PS01180; CUB; 5.  
DR PROSITE; PS00420; SRCR\_1; 8.  
DR PROSITE; PS0287; SRCR\_2; 8.  
DR PROSITE; PS00682; ZP\_DOMAIN; UNKNOWN\_1.  
KW Signal; Transmembrane; Alternative splicing.  
FT SIGNAL 1 28  
FT CHAIN 29 2083  
FT DOMAIN 29 2042  
FT DOMAIN 986 1017  
FT DOMAIN 1250 1259  
FT DOMAIN 1610 1625  
FT DOMAIN 2043 2065  
FT TRANSMEM 2066 2083  
FT DOMAIN 2030 2083  
FT VARSPLIC  
FT CRP-BETA).  
FT THIDRGQPPSTKL -> PELSSQSGSSH (LN  
SEQUENCE 2083 AA; 226734 MW; 6DBE21C7737B4A0A CRC64;

Query Match 48.4%; Score 45; DB 11; Length 2083;  
Best Local Similarity 61.5%; Pred. No. 1.8e+02;  
Matches 8; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 2 YAWRVHARGIRWL 14  
DB 112 YLWRCSHRG--WL 122

RESULT 3  
Q9HA69 PRELIMINARY; PRT; 125 AA.  
ID Q9HA69;  
AC Q9HA69;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE CDNA FLJ12151 FIS, CLONE MAMMAL1000431.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
SEQUENCE FROM N.A.  
RC TISSUE=MAMMARY GLAND;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;  
RT "NEDO human cDNA sequencing project";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK022213; BAB1987.1; -.  
DR SEQUENCE 125 AA; 13762 MW; 37C386B66AC51B49 CRC64;

Query Match 47.3%; Score 44; DB 4; Length 125;  
Best Local Similarity 38.5%; Pred. No. 15;  
Matches 5; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 WRVHARGIRWL 16  
DB 37 WKICHLGFNWIR 49

RESULT 4  
Q9AFZ1 PRELIMINARY; PRT; 133 AA.  
ID Q9AFZ1  
AC Q9AFZ1;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE IS91 ORFB, FRAGMENT.  
GN S0019.  
OS Shigella flexneri.  
OG Plasmid virulence pWR501.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Venkatesan M.N., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,  
RA Blattner F.R.;  
RT "Complete DNA Sequence and Analysis of the Large Virulence Plasmid of  
Shigella flexneri";  
RL Infect Immun. 0:0-0(2001).  
KW EMBL; AF348706; AAK18330.1; -.  
DR Plasmid.  
DR SEQUENCE 133 AA; 15379 MW; 759A64406ACA4F20 CRC64;

Query Match 47.3%; Score 44; DB 2; Length 133;  
Best Local Similarity 37.5%; Pred. No. 16;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 KYAWRVHARGIRWL 16  
DB 112 YLWRCSHRG--WL 122

```

Query Match      47.3%; Score 44; DB 3; Length 481;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      1 KYAWRVAHRGIRWLLR 16
      :| |||.| : |||:
DB      460 QYQPRVAYRHLEWLLK 475

RESULT      7
O80923      PRELIMINARY;      PRT;      125 AA.
ID      O80923
AC      O80923;
DT      O1-NOV-1998 (TrEMBLrel. 08, Created)
DT      O1-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT      O1-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      F13M22.3 PROTEIN.
GN      F13M22.3.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosoids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV, COLUMBIA;
RA      Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
RA      Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
RA      Somerville C.R., Venter J.C.;
RT      "Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.";
RL      Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC      -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
CC      EMBL: AC004684; AAC23624.1; -.
DR      InterPro: IPR000564; 2FE2S_ferredoxin.
DR      PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW      Iron-sulfur.
SQ      SEQUENCE      125 AA;      13953 MW;      0B60805BC5B18C2A CRC64;

```

Query Match	46.2%	Score 43;	DB 10;	Length 125;
Best Local Similarity	47.1%	Pred. No. 21;		
Matches	8;	Conservative	2;	Mismatches 5;
				Indels 2;
				Gaps 1;
QY	1	KYAWRVAAH--RGIRWLL	15	
		: : :		
Db	19	KFAWRIPHVYGATWTL	35	
RESULT	8			
Q9HXU2		PRELIMINARY;		
ID	Q9HXU2	PRT;	183	AA.
AC	Q9HXU2			
DT	01-MAR-2001	(TREMBLrel. 16, Created)		
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)		
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)		
DE	HYPOTHETICAL PROTEIN PA3698.			
GN	PA3698			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=287;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PA01;			
RX	MEDLINE=20437337; PubMed=10984043;			
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,			
RA	Hickey M.J., Brinkman F.S.L., Huftnagle W.O., Kowalik D.J., Lagrou M.,			
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;			
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an			
	opportunistic pathogen."			

Wed Feb 13 07:52:45 2002

SQ SEQUENCE 213 AA; 24384 MW; 3F3AF879C11704A0 CRC64;

RL Nature 406:959-964(2000).

DR EMBL; AE004789; AAG07086.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 183 AA; 20351 MW; 0CF452AB5251FEE2 CRC64;

Query Match 46.2%; Score 43; DB 2; Length 183;  
Best Local Similarity 50.0%; Pred. No. 32;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 YAWRVAHGRIRWLL 15

I : : : : : I

| : : : : |

DB 80 YGFTVSHQGRWYL 93.

QY 8 HGRIRWL 14

I : : : : : I

DB 158 HGRIRWL 164

SQ SEQUENCE 213 AA; 24384 MW; 3F3AF879C11704A0 CRC64;

Query Match 46.2%; Score 43; DB 5; Length 213;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 HGRIRWL 14

I : : : : : I

DB 158 HGRIRWL 164

SQ SEQUENCE 213 AA; 24384 MW; 3F3AF879C11704A0 CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

Query Match 46.2%; Score 43; DB 5; Length 221;

Best Local Similarity 46.7%; Pred. No. 38;

Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 KY--AWRVAHGRIRW 13

I : : : : : I

DB 58 KYGGTWKLRHGRW 72

SQ SEQUENCE 221 AA; 25290 MW; D50FC4DB262DA38D CRC64;

```

GN SPAPB21F2.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Saunders D., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL590562; CAC36890.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 378 AA; 45008 MW; CBAFFE191F8B93E2 CRC64;

Query Match 45.78; Score 42.5; DB 3; Length 378;
Best Local Similarity 57.18; Pred. No. 79;
Matches 8; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 4 WRVAHGI-RWLLR 16
Db 218 WRWHSFNRLIR 231
|| || |||:|
|| || |||:|

RESULT 12
Q9AFQ8 PRELIMINARY; PRT; 133 AA.
AC Q9AFQ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE I591 TRANSPOSASE, FRAGMENT.
GN S0184.
OS Shigella flexneri.
OC Plasmid virulence pWR501.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blattner F.R.;
RT "Complete DNA Sequence and Analysis of the Large Virulence Plasmid of
RT Shigella flexneri.";
RL Infect. Immun. 0:0-0(2001).
DR EMBL: AF348706; AAK18494.1; -
KW Plasmid.
SQ SEQUENCE 133 AA; 15361 MW; 759A64452FDE1B20 CRC64;

Query Match 45.28; Score 42; DB 2; Length 133;
Best Local Similarity 37.58; Pred. No. 33;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KYANVVAHGI-RWLLR 16
Db 78 RFSHVAADKGFVIR 93
:::| || | |::|

RESULT 13
Q9V3U3 PRELIMINARY; PRT; 293 AA.
AC Q9V3U3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG14041 PROTEIN.
GN SP555 OR CG14041.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```

```

OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazcolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue H.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang X.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Serano T.L., Pendleton J.D., Rubin G.M.;
RT "A reverse genetic screen for genes involved in Drosophila eye
RT development.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003608; AAF52206.1; -
DR EMBL: AF231038; AAF34807.1; -
DR FlyBase; FBgn0040280; SP555.
DR InterPro; IPR001870; Gamma_carboxylase.
DR InterPro; IPR001496; SOCS.
DR InterPro; IPR003877; SPRY.
DR InterPro; IPR003878; SPRY_domain.
DR Pfam; PF00622; SPRY.
DR SMART; SM00253; SOCS; 1.
DR SMART; SM00449; SPRY; 1.
SQ SEQUENCE 293 AA; 32902 MW; 6102146A0054D999 CRC64;

Query Match 45.28; Score 42; DB 5; Length 293;
Best Local Similarity 30.88; Pred. No. 73;
Matches 4; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KYANVVAHGI-RW 13
Db 149 EHGWLHSHKGVLM 161
::| ::|::|

RESULT 14

```

Query Match 45.2%; Score 42; DB 2; Length 313;  
 Best Local Similarity 57.1%; Pred. No. 78;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 YARVAHRCIRWLL 15  
 DB 27 FAWRCAERGETLL 40

Search completed: February 12, 2002, 12:38:45  
 Job time: 758 sec

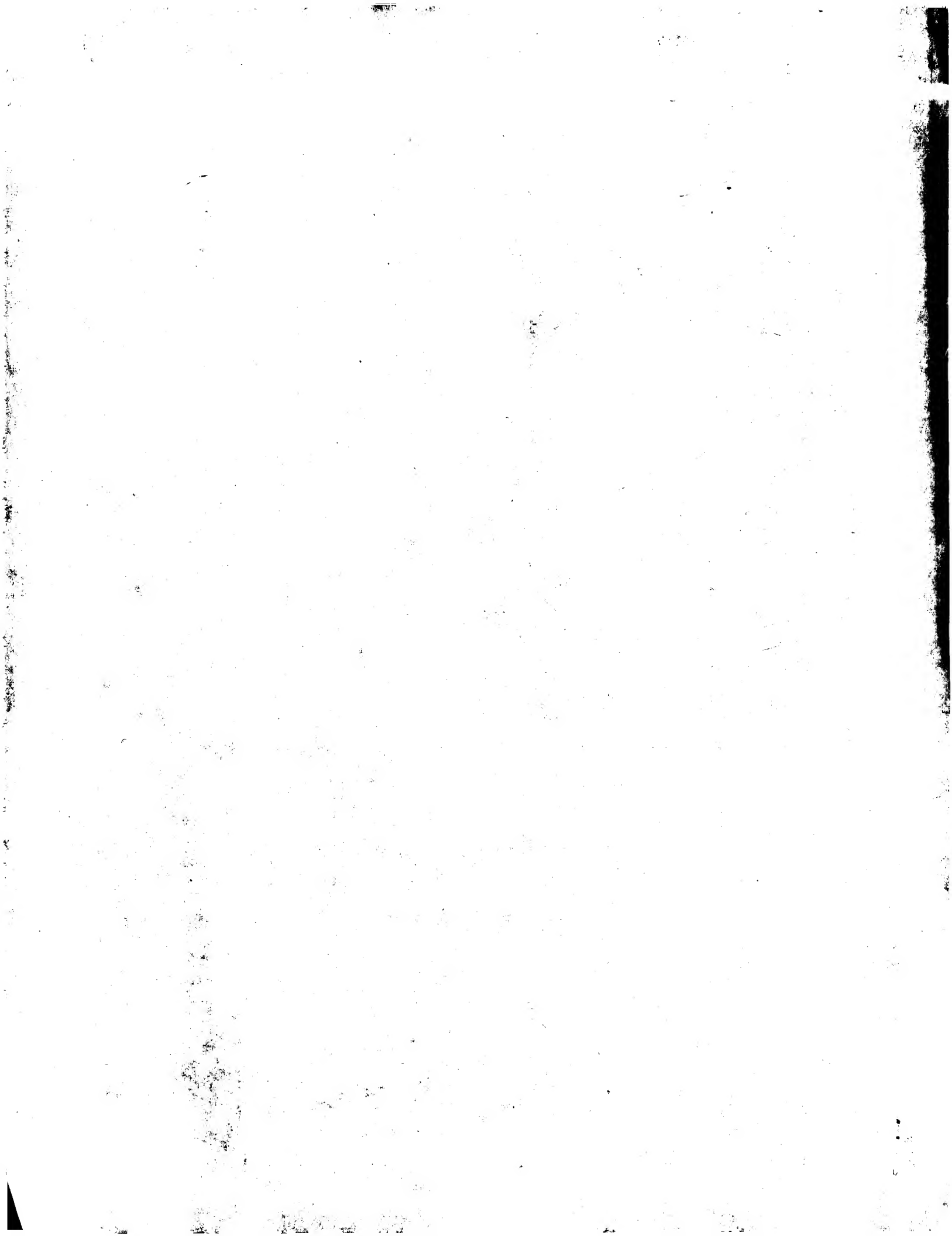
Q9P4W5  
 ID Q9P4W5 PRELIMINARY; PRT; 309 AA.  
 AC Q9P4W5;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE PROTEIN SERINE/THREONINE PHOSPHATASE.  
 GN NORI.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2359/152;  
 RX MEDLINE=20270185; PubMed=10809730;  
 RA Chen X.J., Bauer B.E., Kuchler K., Clark-Walker G.D.;  
 RT "Positive and negative control of multidrug resistance by the Sit4  
 protein phosphatase in Kluyveromyces lactis.";  
 RL J. Biol. Chem. 275:14865-14872(2000).  
 CC -|- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +  
 CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).  
 CC -|- SIMILARITY: TO SERINE/THREONINE SPECIFIC PROTEIN PHOSPHATASE (OR  
 CC PPP) FAMILY.  
 DR EMBL: X87624; CAA60955.1; .  
 DR InterPro: IPR000934; Ser\_thr\_phosphatse.  
 DR Pfam: PF00149; Stphosphatase; 1.  
 DR PRINTS: PR00114; STPHPHATASE.  
 DR SMART: SM00156; PP2AC; 1.  
 DR PROSITE: PS00125; SER\_THR\_PHOSPHATASE; UNKNOWN\_1.  
 KW Hydrolase; Iron; Manganese.  
 SQ SEQUENCE 309 AA; 35308 MW; 1BC8D7FCB0661F37 CRC64;

Query Match 45.2%; Score 42; DB 3; Length 309;  
 Best Local Similarity 58.3%; Pred. No. 77;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 AWRVAHRCIRWL 14  
 DB 205 ANQVSPRGAGWL 216

RESULT 15  
 Q9KXB9  
 ID Q9KXB9 PRELIMINARY; PRT; 313 AA.  
 AC Q9KXB9;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE ARSENICAL PUMP-DRIVING ATPASE.  
 GN BH1795.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001513; BAB05514.1; .  
 DR InterPro: IPR003348; Arsa\_ATPase.  
 DR Pfam: PF02374; Arsa\_ATPase; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 313 AA; 35531 MW; B530B49414F70376 CRC64;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:49 ; Search time 67.2 Seconds  
(without alignments)  
9.821 Million cell updates/sec

Title: US-09-485-571-20

Perfect score: 33

Sequence: 1 XGGXXXXXXXXXXXXXG 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36564827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	16	48.5	347	1	PBX1_MOUSE	P41778 mus musculus
2	16	48.5	394	1	CIW3_HUMAN	O14649 homo sapien
3	16	48.5	430	1	PBX1_HUMAN	P40424 homo sapien
4	16	48.5	1139	1	SRE2_CRIGR	O60429 cricetus
5	15	45.5	110	1	RLA4_YEAST	P02400 saccharomyc
6	15	45.5	306	1	CH38_DROME	P07183 drosophila
7	15	45.5	311	1	BLAC_STRCE	O06650 streptomyce
8	15	45.5	430	1	PBX2_HUMAN	P40425 homo sapien
9	15	45.5	430	1	PBX2_MOUSE	O35984 mus musculus
10	15	45.5	520	1	PAX7_HUMAN	P23759 homo sapien
11	15	45.5	1067	1	SGG_DROME	P18431 drosophila
12	15	45.5	1772	1	MSPL_PLAYO	P13828 plasmodium
13	14	42.4	111	1	RLA2_ASFFU	Q90426 aspergillus
14	14	42.4	162	1	19KD_MYCIT	P31502 mycobacteri
15	14	42.4	172	1	CH18_DROME	P07184 drosophila
16	14	42.4	270	1	BASI_RABIT	Q28740 oryctolagus
17	14	42.4	376	1	EXD_DROME	P40427 drosophila
18	14	42.4	440	1	GAT4_MOUSE	O08369 mus musculus
19	14	42.4	451	1	IFR1_HUMAN	O00458 homo sapien
20	14	42.4	493	1	GATA_RICPR	Q92410 rickettsia
21	14	42.4	515	1	MEF2_DROME	P40791 drosophila
22	14	42.4	537	1	SP70_DICDI	P15269 dictyosteli
23	14	42.4	549	1	DSX_DROME	P23023 drosophila
24	14	42.4	600	1	SP96_DICDI	P14328 dictyosteli
25	14	42.4	747	1	ELS_BOVIN	O04985 bos taurus
26	14	42.4	750	1	ELS_CHICK	P07916 gallus gall
27	14	42.4	833	1	VIRA_AGRT5	P18540 agrobacteri
28	14	42.4	860	1	ELS_MOUSE	P54320 mus musculus
29	14	42.4	1113	1	PER3_MOUSE	O70361 mus musculus
30	14	42.4	1224	1	PER_MOUSE	P07663 drosophila
31	14	42.4	1302	1	NRG_DROME	P20241 drosophila
32	14	42.4	1901	1	YZ08_MYCTU	O53553 mycobacteri
33	14	42.4	3736	1	TRX_DROME	P20659 drosophila

ALIGNMENTS					
RESULT	1				
PBX1_MOUSE					
ID	PBX1_MOUSE	STANDARD;	PRT;	347 AA.	
AC	P41778;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-1 (HOMEBOX PROTEIN PBX1).				
GN	PBX1 OR PBX-1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Adrenal gland;				
RX	MEDLINE=94308119; PubMed=7913464;				
RA	Kagawa N., Ogo A., Takahashi Y., Iwamatsu A., Waterman M.R.;				
RT	"A CAMP-regulatory sequence (CRS1) of Cyp17 is a cellular target for the homeodomain protein Pbx1.";				
RL	J. Biol. Chem. 269:18716-18719(1994).				
CC	-!- FUNCTION: PLAYS A ROLE IN THE CAMP-DEPENDENT REGULATION OF CYP17 GENE EXPRESSION VIA ITS CAMP-REGULATORY SEQUENCE (CRS1)				
CC	5'-ATCAATCAA-3', PROBABLE POSITIVE TRANSCRIPTION REGULATOR. MAY HAVE A ROLE IN STEROIDOGENESIS AND SUBSEQUENTLY SEXUAL DEVELOPMENT AND DIFFERENTIATION.				
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.				
CC	-!- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	-!- TISSUE SPECIFICITY: WIDELY DISTRIBUTED IN STEROIDOGENIC AND NON-STEROIDOGENIC CELLS.				
CC	-!- SIMILARITY: BELONGS TO THE TALE/PBX FAMILY OF HOMEBOX PROTEINS.				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
DR	EMBL; L27453; AAA21832.1; -				
DR	HSSP; P01366; IAKH.				
DR	TRANSFAC; T02088; -				
DR	MGD; MGT:97495; Pbx1.				
DR	InterPro; IPR000047; HTH_repressr.				
DR	InterPro; IPR001356; Homeobox.				
DR	Pfam; PF00046; homeobox; 1				
DR	PRINTS; PR00031; HTHREPRESSR.				
DR	SMART; SM00389; HOX; 1.				
DR	PROSITE; PS00027; HOMEBOX_1; 1.				
DR	PROSITE; PS00071; HOMEBOX_2; 1.				
KW	Transcription regulation; DNA-binding; Nuclear protein; Activator;				
KW	Homeobox; Alternative splicing; Steroidogenesis;				
FT	DOMAIN 127 135				
	POLY-ALA.				

P34527 caenorhabdi  
P02449 dromaius no  
P42037 alternaria  
P05389 drosophila  
Q50418 methylobaci  
P05782 xenopus lae  
O02786 bos taurus  
P13346 mus musculu  
Q12837 homo sapien  
Q63934 mus musculu  
P46152 rattus norv  
Q11067 caenorhabdi

Wed Feb 13 07:51:55 2002

us-09-485-571-20.rsp

FT DNA\_BIND 233 295 HOMEBOX (TALE-TYPE).  
SQ SEQUENCE 347 AA; 38427 MW; C4A2BDD4A410C20 CRC64;

Query Match 48.5%; Score 16; DB 1; Length 347;  
Best Local Similarity 17.6%; Pred. No. 2.5e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 124 GGSAAAAAASGAG 140

RESULT 2  
ID CIW3\_HUMAN STANDARD; PRT; 394 AA.  
AC O14649;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE POTASSIUM CHANNEL SUBFAMILY K MEMBER 3 (ACID-SENSITIVE POTASSIUM CHANNEL PROTEIN TASK) (TWIK-RELATED ACID-SENSITIVE K<sup>+</sup> CHANNEL).  
GN KCNK3 OR TASK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=97459932; PubMed=9312005;  
RA Duprat F., Lesage F., Fink M., Reyes R., Heurteaux C., Lazdunski M.;  
RT TASK, a human background K<sup>+</sup> channel to sense external pH variations  
RT near physiological pH.  
RL EMBO J. 16:5464-5471(1997).  
RN [2]  
RP ACTIVATION.  
RX MEDLINE=99254548; PubMed=10321245;  
RA Patel A.J., Honore E., Lesage F., Fink M., Romey G., Lazdunski M.;  
RT "Inhalational anesthetics activate two-pore-domain background K<sup>+</sup>  
channels.";  
RL Nat. Neurosci. 2:422-426(1999).  
CC -1- FUNCTION: PH-DEPENDENT, VOLTAGE-INSENSITIVE, BACKGROUND POTASSIUM  
CC CHANNEL PROTEIN. RECTIFICATION DIRECTION RESULTS FROM POTASSIUM  
CC ION CONCENTRATION ON EITHER SIDE OF THE MEMBRANE. ACTS AS AN  
CC OUTWARD RECTIFIER WHEN EXTERNAL POTASSIUM CONCENTRATION IS LOW.  
CC WHEN EXTERNAL POTASSIUM CONCENTRATION IS HIGH, CURRENT IS INWARD.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC EXPRESSION IN PANCREAS AND PLACENTA. LOWER EXPRESSION IN BRAIN.  
CC LUNG, PROSTATE, HEART, KIDNEY, UTERUS, SMALL INTESTINE AND COLON.  
CC -1- MISCELLANEOUS: INHIBITED BY EXTERNAL ACIDIFICATION. ACTIVATED BY  
CC HALOTHANE AND ISOFLURANE.  
CC -1- SIMILARITY: BELONGS TO THE TWO PORE DOMAIN FAMILY OF POTASSIUM  
CC CHANNELS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC EMBL: AF006823; AAC51777.1;  
CC MIM: 603220;  
DR InterPro: IPR003280; 2poreK channel.  
DR InterPro: IPR001622; Channel\_pore\_K.  
DR InterPro: IPR003092; TASK channel.  
DR InterPro: IPR000099; TWIK channel.  
DR Pfam: PF02034; TWIK\_channel; 1.  
DR PRINTS: PRO1095; TASKCHANNEL.  
DR PRINTS: PRO1333; 2PORECHANNEL.

KW Ionic channel; Transmembrane; Ion transport; Potassium transport;  
KW Glycoprotein.  
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL);  
FT TRANSMEM 9 29 POTENTIAL.  
FT DOMAIN 78 101 PORE-FORMING (POTENTIAL).  
FT TRANSMEM 108 128 POTENTIAL.  
FT DOMAIN 129 158 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 159 179 POTENTIAL.  
FT DOMAIN 184 207 PORE-FORMING (POTENTIAL).  
FT TRANSMEM 223 243 POTENTIAL.  
FT DOMAIN 244 394 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 53 N-LINKED (GLCNAC... (POTENTIAL).  
SQ SEQUENCE 394 AA; 43518 MW; 9FF4C866F615FB7 CRC64;

Query Match 48.5%; Score 16; DB 1; Length 394;  
Best Local Similarity 17.6%; Pred. No. 2.8e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 2 GGXXXXXXXG 18  
DB 276 GGSATTTDTASSTAAG 292

RESULT 3  
ID PBX1\_HUMAN STANDARD; PRT; 430 AA.  
AC P40424;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-1 (HOMEBOX PROTEIN PBX1)  
DE (HOMEBOX PROTEIN PRL).  
GN PBX1 OR PRL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92049345; PubMed=1682799;  
RA Monica K., Saltman D., Nourse J., Gallili N., Cleary M.L.;  
RT "PBX2 and PBX3, new homeobox genes with extensive homology to the  
human proto-oncogene PBX1.";  
RL Mol. Cell. Biol. 11:6149-6157(1991).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM PBX1A).  
RX MEDLINE=21167395; PubMed=11267683;  
RA Thameem F., Wolford J.K., Bogardus C., Prochazka M.;  
RT "Analysis of PBX1 as a candidate gene for type 2 diabetes mellitus in  
pima Indians.";  
RL Biochim. Biophys. Acta 1518:215-220(2001).  
RN [3]  
RP SEQUENCE OF 89-430 FROM N.A., AND CHROMOSOMAL TRANSLOCATION.  
RX MEDLINE=90150282; PubMed=1967983;  
RA Kamps M.P., Murie C., Sun X.-H., Baltimore D.;  
RT "A new homeobox gene contributes the DNA binding domain of the t(1;19)  
translocation protein in pre-B ALL.";  
RL Cell 60:547-555(1990).  
RN [4]  
RP CHROMOSOMAL TRANSLOCATION.  
RX MEDLINE=91129319; PubMed=1671560;  
RA Hunger S.P., Gallili N., Carroll A.J., Crist W.M., Link M.P.,  
RA Cleary M.L.;  
RT "The t(1;19)(q23;p13) results in consistent fusion of E2A and PBX1  
coding sequences in acute lymphoblastic leukemias.";  
RL Blood 77:687-693(1991).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=93317624; PubMed=8327485;  
RA van Dijk M.A., Voorhoeve P.M., Murie C.;  
RT "PBX1 is converted into a transcriptional activator upon acquiring  
the N-terminal region of E2A in pre-B-cell acute lymphoblastoid

leukemia.";  
 [6]  
 CHARACTERIZATION.  
 RX MEDLINE-94254851; PubMed-7910944;  
 RA Lu Q., Wright D.D., Kamps M.P.;  
 RT "Fusion with E2A converts the Pbx1 homeodomain protein into a  
 RT constitutive transcriptional activator in human leukemias carrying  
 RT the t(1;19) translocation.";  
 RL Mol. Cell. Biol. 14:3938-3948(1994).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 233-319 IN COMPLEX TO HOXB1.  
 RX MEDLINE-99159825; PubMed-10052460;  
 RA Piper D.E., Batchelor A.H., Chang C.-P., Cleary M.L., Wolberger C.;  
 RT "Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the  
 RT hexapeptide and a fourth homeodomain helix in complex formation.";  
 RL Cell 96:587-597(1999).  
 CC -1- FUNCTION: THIS NONACTIVATING PROTEIN WHICH COULD BE A REPRESSOR;  
 CC BINDS THE SEQUENCE 5'-ATCAATCA-3'. IT IS CONVERTED INTO A POTENT  
 CC TRANSCRIPTIONAL ACTIVATOR BY THE (1;19) TRANSLOCATION. MAY HAVE A  
 CC ROLE IN STEROIDOGENESIS AND SUBSEQUENTLY SEXUAL DEVELOPMENT AND  
 CC DIFFERENTIATION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PBX1A (SHOWN HERE) AND PBX1B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXCEPT IN CELLS OF  
 CC THE B AND T LINEAGE.  
 CC -1- DISEASE: A FORM OF PRE-B-CELL ACUTE LYMPHOBLASTIC LEUKEMIA (B-ALL)  
 CC (25% OF HUMAN PEDIATRIC CASES) IS CHARACTERIZED BY A CHROMOSOMAL  
 CC TRANSLOCATION T(1;19)(Q23;P13.3) WHICH INVOLVES PBX1 AND E2A  
 CC GENES. E2A-PBX1 TRANSFORMS CELLS BY CONSTITUTIVELY ACTIVATING  
 CC TRANSCRIPTION OF GENES REGULATED BY PBX1 OR BY OTHER MEMBERS OF  
 CC THE PBX PROTEIN FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE TALE/PBX FAMILY OF HOMEOBOX PROTEINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M86546; AAA60031.1; -;  
 CC EMBL; AF313404; AAG30941.1; -;  
 CC EMBL; AF313396; AAG30941.1; JOINED.  
 CC EMBL; AF313397; AAG30941.1; JOINED.  
 CC EMBL; AF313398; AAG30941.1; JOINED.  
 CC EMBL; AF313399; AAG30941.1; JOINED.  
 CC EMBL; AF313400; AAG30941.1; JOINED.  
 CC EMBL; AF313401; AAG30941.1; JOINED.  
 CC EMBL; AF313402; AAG30941.1; JOINED.  
 CC EMBL; AF313403; AAG30941.1; JOINED.  
 CC EMBL; M31522; AAA36764.1; JOINED.  
 CC PDB; 1B72; 19-FEB-99.  
 CC TRANSFAC; T01481; -;  
 CC MIM; 176310; -;  
 CC InterPro; IPR000047; HTH\_repressor.  
 CC InterPro; IPR001356; Homeobox.  
 CC Pfam; PF000046; homeobox; 1.  
 CC PRINTS; PR00031; HTHREPRESSR.  
 CC SMART; SM00389; HOX; 1.  
 CC PROSITE; PS00027; HOMEOBOX 1; 1.  
 CC PROSITE; PS50071; HOMEOBOX 2; 1.  
 CC Transcription regulation; DNA-binding; Nuclear protein; Activator;  
 KW Repressor; Homeobox; Proto-oncogene; Chromosomal translocation;  
 KW Alternative splicing; Steroidogenesis; Sexual differentiation;  
 KW 3D-structure.  
 FT DOMAIN 127 135 POLY-ALA.  
 FT DNA\_BIND 233 295  
 FT SITE 88 89 BREAKPOINT FOR TRANSLOCATION TO FORM  
 FT CAA67062A-PBX1 ONCOGENE.  
 FT VARSPLIC 334 347 SSSSFNMSNGDLF -> GYSPCYQDPRRIQ (IN

FT VARSPLIC 348 430 ISOFORM PBX1B).  
 FT MISSING (IN ISOFORM PBX1B).  
 SQ SEQUENCE 430 AA; 45626 MW; AD3FFACBC5A9E715 CRC64;  
 Query Match 48.5%; Score 16; DB 1; Length 430;  
 Best Local Similarity 17.6%; Pred. No. 3e+03;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 2 GGXXXXXXXGXXXXXG 18  
 DB 124 GGSAAAAAASGAG 140  
 RESULT 4  
 SRE2\_CRIGR STANDARD; PRT; 1139 AA.  
 ID SRE2\_CRIGR Q60418; Q60428; Q60427;  
 AC Q60429; Q60418; Q60428; Q60427;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE STEROL REGULATORY ELEMENT BINDING PROTEIN-2 (SREBP-2) (STEROL  
 DE REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 2).  
 GN SREBP2 OR SREBP2.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetus.  
 OX NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND TRUNCATED FORM SRD-1.  
 RX MEDLINE-95047343; PubMed-7958866;  
 RA Yang J., Sato R., Goldstein J.L., Brown M.S.;  
 RT "Sterol-resistant transcription in CHO cells caused by gene  
 RT rearrangement that truncates SREBP-2.";  
 RL Genes Dev. 8:1910-1919(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A. (TRUNCATED FORMS SRD-1 TO SRD-3).  
 RX MEDLINE-95263566; PubMed-7744865;  
 RA Yang J., Brown M.S., Ho Y.K., Goldstein J.L.;  
 RT "Three different rearrangements in a single intron truncate sterol  
 RT regulatory element binding protein-2 and produce sterol-resistant  
 RT phenotype in three cell lines. Role of introns in protein  
 RT evolution.";  
 RL J. Biol. Chem. 270:12152-12161(1995).  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO THE STEROL  
 CC REGULATORY ELEMENT 1 (SRE-1) (5'-ATCACCAC-3') FOUND IN THE  
 CC FLANKING REGION OF THE LDLR AND HMG-COA SYNTHASE GENES (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANES OF THE NUCLEAR  
 CC ENVELOPE AND ENDOPLASMIC RETICULUM. RELEASED INTO THE NUCLEUS  
 CC UPON PROTEOLYTIC CLEAVAGE.  
 CC -1- PTM: UNDER STEROL-DEPLETED CONDITIONS, SREBPS ARE PROTEOLYTICALLY  
 CC CLEAVED TO PRODUCE N-TERMINAL FRAGMENTS WHICH ENTER THE NUCLEUS  
 CC AND ACTIVATE TRANSCRIPTION. SIMILAR CLEAVAGE BY THE CYSTEINE  
 CC PROTEASES, APOPAIN AND CASPASE-7, IS INDUCED DURING APOPTOSIS,  
 CC INDEPENDENT OF STEROL LEVELS.  
 CC -1- DISEASE: STEROL-RESISTANT DEFECTIVE (SRD) PHENOTYPES EXPRESS  
 CC TRUNCATED FORMS OF SREBP-2 PROTEIN, OFTEN FOUND FUSED TO OTHER  
 CC PROTEINS, AS IS THE CASE IN SRD-1, WHERE SREBP-2 IS FUSED TO AN  
 CC OUT-OF-FRAME KU P70 PROTEIN OR, IN SRD-2 WHERE THE FUSION PROTEIN  
 CC IS A LIM DOMAIN-CONTAINING PROTEIN. SRD PHENOTYPES ARE RESISTANT  
 CC TO STEROL BIOSYNTHESIS REPRESSION BY STEROLS.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC  EMBL; U12330; AAA74141.1; -
CC  EMBL; U12329; AAA74140.1; ALT_TERM.
CC  EMBL; U22819; AAA85719.1; ALT_TERM.
CC  EMBL; U22818; AAA85718.1; ALT_TERM.
CC  HSSP; P36956; IAN9.
CC  InterPro; IPR003015; HLH_Myc.
CC  InterPro; IPR001092; HLH_dlm.
CC  Pfam; PF00010; HLH; 1.
CC  SMART; SM00353; HLH; 1.
CC  PROSITE; PS00038; HELIX_LOOP_HELIX; FALSE_NEG.
KW  Transcription regulation; DNA-binding; Nuclear protein; Transmembrane;
KW  Endoplasmic reticulum; Polymorphism; Chromosomal translocation.
FT  DOMAIN 1 479
FT  TRANSMEM 480 500
FT  DOMAIN 501 531
FT  TRANSMEM 532 552
FT  TRANSMEM 553 1139
FT  DOMAIN 1 50
FT  DOMAIN 52 124
FT  DOMAIN 125 244
FT  BINDING 95 421
FT  DNA_BIND 328 341
FT  DOMAIN 342 379
FT  DOMAIN 378 399
FT  DOMAIN 589 593
FT  DOMAIN 857 860
FT  SITE 460 461
FT  SITE 466 467
FT  VARIANT 493 493
FT  SEQUENCE 1139 AA; 123655 MW; E81C2778EBF02653 CRC64;
SQ
Query Match 48.5%; Score 16; DB 1; Length 1139;
Best Local Similarity 17.6%; Pred. No. 6.2e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXGXXXXXXXXXG 18
DB 57 GGSSGSSSSSSNSSSSSG 73

RESULT 5
RLA4_YEAST STANDARD; PRT; 110 AA.
AC P02400;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1989 (Rel. 11, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 60S ACIDIC RIBOSOMAL PROTEIN P2-BETA (L45) (YL44C) (YPA1) (L12E1A).
GN RPP2B OR RPLA4 OR YL2E1A OR RPL45 OR YDR382W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88243786; PubMed=2837476;
RA Remacha M., Saenz-Robles M.T., Vilella M.D., Ballesta J.P.G.;
RT "Independent genes coding for three acidic proteins of the large
RT ribosomal subunit from Saccharomyces cerevisiae.";
RL J. Biol. Chem. 263:9094-9101(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=SR26-12C;
RC STRAIN=SR26-12C;
RD MEDLINE=90130289; PubMed=2404943;
RA Newton C.H., Shimmin L.C., Yee J., Dennis P.P.;

```

```

RT "A family of genes encode the multiple forms of the Saccharomyces
RT cerevisiae ribosomal proteins equivalent to the Escherichia coli L12
RT protein and a single form of the L10-equivalent ribosomal protein.";
RL J. Bacteriol. 172:579-588(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE.
RX MEDLINE=82069169; PubMed=7030402;
RA Itoh T.;
RT "Primary structure of an acidic ribosomal protein YPA1 from
RT Saccharomyces cerevisiae. Isolation and characterization of peptides
RT and the complete amino acid sequence.";
RL Biochim. Biophys. Acta 671:16-24(1981).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -1- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC -1- MISCELLANEOUS: YEASTS CONTAIN 4 INDIVIDUAL SMALL RIBOSOMAL A
CC PROTEINS (RPA) WHICH CAN BE CLASSIFIED INTO TWO COUPLES OF SIMILAR
CC BUT NOT IDENTICAL SEQUENCES. EACH COUPLE IS DISTINCTLY RELATED TO
CC ONE OF THE TWO A PROTEINS PRESENT IN MULTICELLULAR ORGANISMS.
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; J03761; AAA34972.1; -
CC EMBL; M26505; AAA34732.1; -
CC EMBL; U28373; AAB64818.1; -
CC EMBL; U32274; AAB64824.1; -
CC PIR; A35109; R5BYA1.
CC SGD; S0002790; RPP2B.
CC InterPro; IPR001813; 60s_Ribosomal.
CC Pfam; PF00428; 60s_Ribosomal; 1.
KW Ribosomal protein; Phosphorylation; Multigene family.
FT CONFLICT 75 78 AAGA -> GPAS (IN REF. 4).
FT CONFLICT 86 87 DA -> GD (IN REF. 4).
FT CONFLICT 89 89 E -> A (IN REF. 4).
FT SEQUENCE 110 AA; 11050 MW; EC45406CB5F199F4 CRC64;

Query Match 45.5%; Score 15; DB 1; Length 110;
Best Local Similarity 17.6%; Pred. No. 1.8e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXGXXXXXXXXXG 18
DB 69 GGASSAAAGAGAGAAAGG 85

RESULT 6
CH38_DROME STANDARD; PRT; 306 AA.
ID CH38_DROME
AC P07183; Q9W3E5;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

```

DE CHORION PROTEIN S38.  
GN CP38 OR S38 OR CG1121.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Phylorhoda; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=87246506; PubMed=3036489;  
RA Spradling A.C., de Cicco D.V., Wakimoto B.T., Levine J.F.,  
RA Kafayan L.J., Cooley L.;  
RT Amplification of the X-linked Drosophila chorion gene cluster  
RL requires a region upstream from the s38 chorion gene.";  
RL EMBO J. 6:1045-1053(1987).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=BERKELEY;  
RC MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Fertaz C., Ferreira S., Fleischmann W.,  
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Gloddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X05245; CAA28871.1; -;  
DR EMBL; AE003444; AAF46383.1; -;  
DR PIR; S08607; S08607.  
DR HSSP; P04002; IATF.  
DR FlyBase; FBgn0000360; Cp38.  
KW Chorion.  
SQ SEQUENCE 306 AA; 30448 MW; 2F51C96F9F82DF83 CRC64;

Query Match 45.5%; Score 15; DB 1; Length 306;  
Best Local Similarity 17.6%; Pred. No. 3.9e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Caps 0;  
QY 2 CGXXXXXXXG 18  
DB 39 GGADAASAAAAAGGAG 55  
RESULT 7  
BLAC\_STRCE STANDARD; PRT; 311 AA.  
AC 006650;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE BETA-LACTAMASE PRECURSOR (EC 3.5.2.6) (PENICILLINASE).  
GN BLA.  
OS Streptomyces cellulosae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1968;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KCC S0127; PubMed=7916705;  
RX MEDLINE=93178958;  
RA Ogawara H.;  
RT "Sequence of a gene encoding beta-lactamase from Streptomyces  
RT cellulosae.";  
RL Gene 124:111-114(1993).  
CC -1- FUNCTION: HYDROLYZES BENZYL-PENICILLIN AND CLOXACILLIN (AT 10% OF  
CC THE RATE OF BENZYL-PENICILLIN).  
CC -1- CATALYTIC ACTIVITY: A BETA-LACTAM + H2O -> A SUBSTITUTED BETA-  
CC AMINO ACID.  
CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY. THIS IS  
CC CONTRARY TO THE RESULT EXPECTED FROM ITS SUBSTRATE SPECIFICITY AND  
CC ITS PROPERTY OF BINDING BLUE DEXTRAN AND NADP+.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D12653; BAA02176.1; -;  
DR PIR; JN0520; JN0520.  
DR HSSP; P00810; ITEM.  
DR InterPro; IPR001466; Beta\_lactam.  
DR InterPro; IPR000871; Beta\_lactam.  
DR Pfam; PF00144; beta-lactamase; 1.  
DR PRINTS; P00118; BETA-LACTAMASE.  
DR PROSITE; PS00146; BETA-LACTAMASE A; 1.  
KW Hydrolase; Antibiotic resistance; NADP; Signal.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 34 311 BETA-LACTAMASE.  
FT ACT\_SITE 86 86 BY SIMILARITY.  
FT BINDING 252 254 SUBSTRATE (BY SIMILARITY).  
SQ SEQUENCE 311 AA; 33136 MW; F3578EBEEA92A3FB CRC64;  
Query Match 45.5%; Score 15; DB 1; Length 311;  
Best Local Similarity 17.6%; Pred. No. 3.9e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Caps 0;  
QY 2 GGXXXXXXXG 18  
DB 21 GGALAGSTTASASAG 37

```

DR InterPro; IPR001356; Homeobox.
PFAM: PF00046; homeobox; 1.
DR PRINTS: PR00031; HTHREPRESSR.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-2 (HOMEBOX PROTEIN PBX2)
DE (G17).
DE (G17).
GN PBX2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE-92049345; PubMed-1682799;
RA Monica K., Galili N., Nourse J., Saltman D., Cleary M.L.;
RT "PBX2 and PBX3, new homeobox genes with extensive homology to the
RT human proto-oncogene PBX1."
RT Mol. Cell. Biol. 11:6149-6157(1991).
[2]
SEQUENCE FROM N.A.
RX MEDLINE-95278934; PubMed-7759099;
RA Aguado B., Campbell R.D.;
RT "The novel gene G17, located in the human major histocompatibility
RT complex, encodes PBX2, a homeodomain-containing protein."
RT Genomics 25:650-659(1995).
[3]
SEQUENCE FROM N.A.
RX MEDLINE-95137587; PubMed-7835890;
RA Sugaya K., Fukagawa T., Matsumoto K., Mita K., Takahashi E., Ando A.,
RA Inoko H., Kemura T.;
RT "Three genes in the human MHC class III region near the junction with
RT the class II: gene for receptor of advanced glycosylation end
RT products, PBX2 homeobox gene and a notch homolog, human counterpart
RT of mouse mammary tumor gene int-3."
RT Genomics 23:408-419(1994).
[4]
SEQUENCE FROM N.A.
RX Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banta A., Spies T., Hood L.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
[5]
CHARACTERIZATION.
RX MEDLINE-94254851; PubMed-7910944;
RA Lu Q., Wright D.D., Kamps M.P.;
RT "Fusion with E2A converts the Pbx1 homeodomain protein into a
RT constitutive transcriptional activator in human leukemias carrying
RT the t(1;19) translocation."
RL Mol. Cell. Biol. 14:3938-3948(1994).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS THE SEQUENCE 5'-
CC ATCAATCAA-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE TALE/PBX FAMILY OF HOMEBOX PROTEINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AF020198; AAB71193.1; -.
CC EMBL; AF030001; AAB82006.1; -.
CC HSSP; P01366; IAKH.
CC MGD; MGI:1341793; Pbx2.
DR InterPro; IPR000047; HTH_repressr.
DR InterPro; IPR001356; Homeobox.
PFAM: PF00046; homeobox; 1.
DR PRINTS; PR00031; HTHREPRESSR.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator;
KW Homeobox.
FT DNA_BIND 244 306 HOMEBOX (TALE-TYPE).
FT DOMAIN 8 11 POLY-PRO.
InterPro; IPR000047; HTH_repressr.

```

```

DR InterPro; IPR001356; Homeobox.
PFAM: PF00046; homeobox; 1.
DR PRINTS: PR00031; HTHREPRESSR.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator;
KW Homeobox.
FT DNA_BIND 244 306 HOMEBOX (TALE-TYPE).
FT DOMAIN 8 11 POLY-PRO.
SQ SEQUENCE 430 AA; 45881 MW; EF2FFA158C4DAF68 CRC64;

Query Match 45.5%; Score 15; DB 1; Length 430;
Best Local Similarity 17.6%; Pred. No. 5e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXXXXXXXG 18
   ||
Db 133 GGGSAAAAAAAAAAASGGG 149

RESULT 9
PBX2_MOUSE STANDARD; PRT; 430 AA.
ID ID PBX2_MOUSE
AC O35984;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-2 (HOMEBOX PROTEIN PBX2).
GN PBX2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RL Liu Y., Macdonald R.J.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RX Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RA Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS THE SEQUENCE 5'-
CC ATCAATCAA-3' (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE TALE/PBX FAMILY OF HOMEBOX PROTEINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AF020198; AAB71193.1; -.
CC EMBL; AF030001; AAB82006.1; -.
CC HSSP; P01366; IAKH.
CC MGD; MGI:1341793; Pbx2.
DR InterPro; IPR000047; HTH_repressr.
DR InterPro; IPR001356; Homeobox.
PFAM: PF00046; homeobox; 1.
DR PRINTS; PR00031; HTHREPRESSR.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator;
KW Homeobox.
FT DNA_BIND 244 306 HOMEBOX (TALE-TYPE).
FT DOMAIN 8 11 POLY-PRO.
InterPro; IPR000047; HTH_repressr.

```



FT DOMAIN 137 145 POLY-ALA.  
SQ SEQUENCE 430 AA; 45809 MW; CB6B71A6FE207EBD CRC64;

Query Match 45.5%; Score 15; DB 1; Length 430;  
Best Local Similarity 17.6%; Pred. No. 5e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXGXXXXXG 18

Db 133 GGGSAASAAAAASGGG 149

RESULT 10

PAX7\_HUMAN

ID PAX7\_HUMAN STANDARD; PRT; 520 AA.

AC P23759;

DT 01-NOV-1991 (Rel. 20, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PAIRED BOX PROTEIN PAX-7 (HUP1).

GN PAX7 OR HUP1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RX MEDLINE=97480728; PubMed=9333373;

RA Vorobyov E., Mertsalov I., Dockhorn-Dworniczak B., Dworniczak B.,

RA Horst J.;

RT "The genomic organization and full coding region of the human PAX7

gene.";

RL Genomics 45:168-174(1997).

RN [2]

RP SEQUENCE FROM N.A. (LONG FORM).

RA Heath P.;

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-467 FROM N.A. (LONG FORM).

RX MEDLINE=95075634; PubMed=7521137;

RA Schaefer B.W., Czerny T., Bernasconi M., Genini M., Busslinger M.;

RT "Molecular cloning and characterization of a human PAX-7 cDNA

expressed in normal and neoplastic myocytes.";

RL Nucleic Acids Res. 22:4574-4582(1994).

RN [4]

RP SEQUENCE OF 30-195 FROM N.A. (SHORT FORM).

RX MEDLINE=89305521; PubMed=2501086;

RA Burri M., Tromvoukis Y., Bopp D., Frigerio G., Noll M.;

RT "Conservation of the paired domain in metazoans and its structure in

three isolated human genes.";

RL EMBO J. 8:1183-1190(1989).

CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. IT MAY HAVE A ROLE IN

CC MYOGENESIS.

CC -1- SUBUNIT: CAN BIND AS A HETERODIMER WITH PAX3.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND

CC A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- DISEASE: RHABDOMYOSARCOMA-2 (RMS2) IS CHARACTERIZED BY A

CC CHROMOSOMAL TRANSLOCATION T(1;13)(P36;Q14) WHICH INVOLVES PAX7 AND

CC FOXO1A. THE RESULTING PROTEIN IS A TRANSCRIPTIONAL ACTIVATOR.

CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.

CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X96743; CAA65520.1; -.

DR

DR EMBL; X96744; CAA65521.1; -.  
DR EMBL; X15042; CAA65521.1; JOINED.  
DR EMBL; X15250; CAA65521.1; JOINED.  
DR EMBL; X15251; CAA65521.1; JOINED.  
DR EMBL; X96745; CAA65521.1; JOINED.  
DR EMBL; X96746; CAA65521.1; JOINED.  
DR EMBL; X96747; CAA65521.1; JOINED.  
DR EMBL; X96748; CAA65521.1; JOINED.  
DR EMBL; X96744; CAA65522.1; -.  
DR EMBL; X15042; CAA65522.1; JOINED.  
DR EMBL; X15250; CAA65522.1; JOINED.  
DR EMBL; X15251; CAA65522.1; JOINED.  
DR EMBL; X96745; CAA65522.1; JOINED.  
DR EMBL; X96746; CAA65522.1; JOINED.  
DR EMBL; X96747; CAA65522.1; JOINED.  
DR EMBL; X96748; CAA65522.1; JOINED.  
DR EMBL; AL021528; CAA16432.1; -.  
DR EMBL; Z35141; CAA84513.1; -.  
DR PIR; S06959; S06959.  
DR HSP; P06601; IFJL.  
DR TRANSFAC; T00396; -.  
DR MIM; 167410; -.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR001523; Paired\_box.  
DR Pfam; PF00046; homeobox; 1.  
DR Pfam; PF00292; PAX; 1.  
DR PRINTS; PR00027; PAIREDBOX.  
DR SMART; SM00389; HOX; 1.  
DR SMART; SM00351; PAX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
DR PROSITE; PS00034; PAIRED\_BOX; 1.  
KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;  
KW Developmental protein; Paired box; Chromosomal translocation;  
KW Proto-oncogene; Alternative splicing.  
FT DOMAIN 34 161 PAIRED BOX.  
FT DNA\_BIND 217 276 HOMEBOX.  
FT DOMAIN 340 346 POLY-ALA.  
FT VARSPLIC 151 152 MISSING (IN SHORT ISOFORM).  
SQ SEQUENCE 520 AA; 56896 MW; 3B0F8CC99D65699C CRC64;

Query Match 45.5%; Score 15; DB 1; Length 520;  
Best Local Similarity 17.6%; Pred. No. 5.8e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXGXXXXXG 18

Db 337 GGLAASAAADTSSAYG 353

RESULT 11

SGG\_DROME

ID SGG\_DROME STANDARD; PRT; 1067 AA.

AC P18431; P23646; Q27603; Q27605; Q76881; Q9U094; Q9W4X3; Q27604;

AC Q9NF42;

DT 01-NOV-1990 (Rel. 16, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROTEIN KINASE SHAGGY (EC 2.7.1.-) (PROTEIN ZESTE-WHITE 3).

GN SGG OR ZW3 OR EG:BACR7C10.8 OR EG:155E2.3 OR CG2621.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC TISSUE=Embryo;

RX MEDLINE=93223707; PubMed=8467811;

RA Ruel L., Pantesco V., Lutz Y., Simpson P., Bourouls M.;

RT "Functional significance of a family of protein kinases encoded at the

RT shaggy locus in Drosophila.";

Wed Feb 13 07:51:55 2002

EMBO J. 12:1657-1669(1993).

[2] SEQUENCE FROM N.A. (ISOFORM ZYGOTIC), AND CHARACTERIZATION.  
 RP STRAIN-DP CN BW; TISSUE=Embryo;  
 RX MEDLINE=90361000; PubMed=2118107;  
 RA Bourouls M., Moore P., Ruel L., Grau Y., Heltzler P., Simpson P.;  
 RT "An early embryonic product of the gene shaggy encodes a  
 serine/threonine protein kinase related to the CDC28/cdc2+  
 subfamily.";   
 RL EMBO J. 9:2877-2884(1990).

[3] SEQUENCE OF 193-1067 FROM N.A. (ISOFORMS SGG46 AND ZYGOTIC).  
 RP TISSUE=Embryo, and Ovary;  
 RX MEDLINE=90294930; PubMed=2113617;  
 RA Siegfried E., Perkins L.A., Capaci T.M., Perrimon N.;  
 RT "Putative protein kinase product of the Drosophila segment-polarity  
 gene zeste-white3.";   
 RL Nature 345:825-829(1990).

[4] SEQUENCE FROM N.A.  
 RP STRAIN=OREGON-R;  
 RX MEDLINE=20196011; PubMed=10731137;  
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,  
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,  
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo S.,  
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,  
 RA Beinert N., Dowe G., Schaefer U., Jaecle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
 RA Glover D.M.;   
 RL "From sequence to chromosome: the tip of the X chromosome of D.  
 melanogaster.";   
 RL Science 287:2220-2222(2000).

[5] SEQUENCE FROM N.A.  
 RP STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.C., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazer E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA de Pablo B., Delcher A., Downes M., Dugan-Rocha S., Fleischmann W.,  
 RA Dodson K., Doup L.E., Downes M., Ferraz C., Ferrera S., Glasser K.,  
 RA Durbin K.J., Evangelista C.C., Garq N.S., Gelbart W.M., Glasser K.,  
 RA Foslter C., Gabrielson A.E., Garq N.S., Guan P., Harris M.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Hernandez J.R., Houck J.,  
 RA Harris N.L., Harvey D., Helman T.J., Henshaw J., Ibegwam C.,  
 RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Jalali M., Kalush F., Karpen G.H., Kravitz S., Kulp D., Lai Z.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Li J., Li Z., Liang Y., Lin X.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of Drosophila melanogaster.";   
 RL Science 287:2185-2195(2000).

[6] PHOSPHORYLATION OF TYR-767.  
 RP MEDLINE=93178457; PubMed=8382613;  
 RX Hughes K., Nikolakaki E., Plyte S.E., Totty N.F., Woodgett J.R.;  
 RT "Modulation of the glycogen synthase kinase-3 family by tyrosine  
 phosphorylation.";   
 RL EMBO J. 12:803-808(1993).

[7] PHOSPHORYLATION OF ARM.  
 RP MEDLINE=95113174; PubMed=7529201;  
 RX Peifer M., Pai L.-M., Casey M.;  
 RT "Phosphorylation of the Drosophila adherens junction protein  
 Armadillo: roles for wingless signal and zeste-white 3 kinase.";   
 RL Dev. Biol. 166:543-556(1994).

[8] PROTEIN INTERACTION WITH WG AND EN.  
 RP TISSUE=Embryo;  
 RX MEDLINE=93113685; PubMed=1335365;  
 RA Siegfried E., Chou T.B., Perrimon N.;  
 RT "Wingless signaling acts through zeste-white 3, the Drosophila homolog  
 of glycogen synthase kinase-3, to regulate engrailed and establish  
 cell fate.";   
 RL Cell 71:1167-1179(1992).

-1- FUNCTION: REQUIRED FOR SEVERAL DEVELOPMENTAL EVENTS SUCH AS  
 SYNCTIAL BLASTODERM FORMATION AND EMBRYONIC SEGMENTATION. IS  
 INVOLVED IN TRANSCRIPTIONAL REGULATION. SGG IS REQUIRED FOR ARM  
 PHOSPHORYLATION. WG SIGNALING OPERATES BY INACTIVATING THE SGG  
 REPRESSION OF EN AUTOACTIVATION.  
 -1- CATALYTIC ACTIVITY: ATP + A PROTEIN -> ADP + A PHOSPHOPROTEIN.  
 -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; ZYGOTIC, SGG39 AND  
 SGG46/MATERNAL (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 -1- TISSUE SPECIFICITY: EXPRESSION IS OVER ALL THE EMBRYO AT ALL  
 STAGES, NO LOCAL ACCUMULATION IS OBSERVED.  
 -1- DEVELOPMENTAL STAGE: ISOFORM SGG46 IS EXPRESSED IN 12-24HR EMBRYOS  
 AND PRESENT THROUGHOUT THE LARVAL, PUPAL AND ADULT STAGES. ISOFORM  
 ZYGOTIC IS HIGHLY EXPRESSED IN 0-2HR EMBRYOS AND PRESENT BUT  
 REDUCED THROUGHOUT LATER EMBRYONIC DEVELOPMENT. EXPRESSION  
 PERSISTS THROUGHOUT LARVAL STAGES. ISOFORM SGG39 IS EXPRESSED IN  
 12-24HR EMBRYOS AND PRESENT THROUGHOUT THE LARVAL, PUPAL AND ADULT  
 STAGES.

-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CDC2/CDKX SUBFAMILY; GSK-3 SUBSUBFAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 EMBL: X70862; CAA50212.1; -  
 EMBL: X70863; CAA50213.1; -  
 EMBL: X70864; CAA50214.1; -  
 EMBL: X70865; CAA50215.1; -  
 EMBL: X53332; CAA37419.1; -  
 EMBL: X54005; CAA37951.1; -  
 EMBL: X54006; CAA37952.1; -  
 EMBL: AL034544; CAA22520.1; -  
 EMBL: AL024485; CAA22520.1; JOINED.  
 EMBL: AL024485; CAA19676.1; -  
 EMBL: AL034544; CAA19676.1; JOINED.  
 EMBL: AL121804; CAB65860.1; -  
 EMBL: AL024485; CAB65860.1; JOINED.  
 EMBL: AL121804; CAB72296.1; -  
 EMBL: AL024485; CAB72296.1; JOINED.  
 EMBL: AE003425; AAF45801.1; ALT\_SEQ.  
 PIR: S10931; S10931.  
 PIR: S11675; S11675.

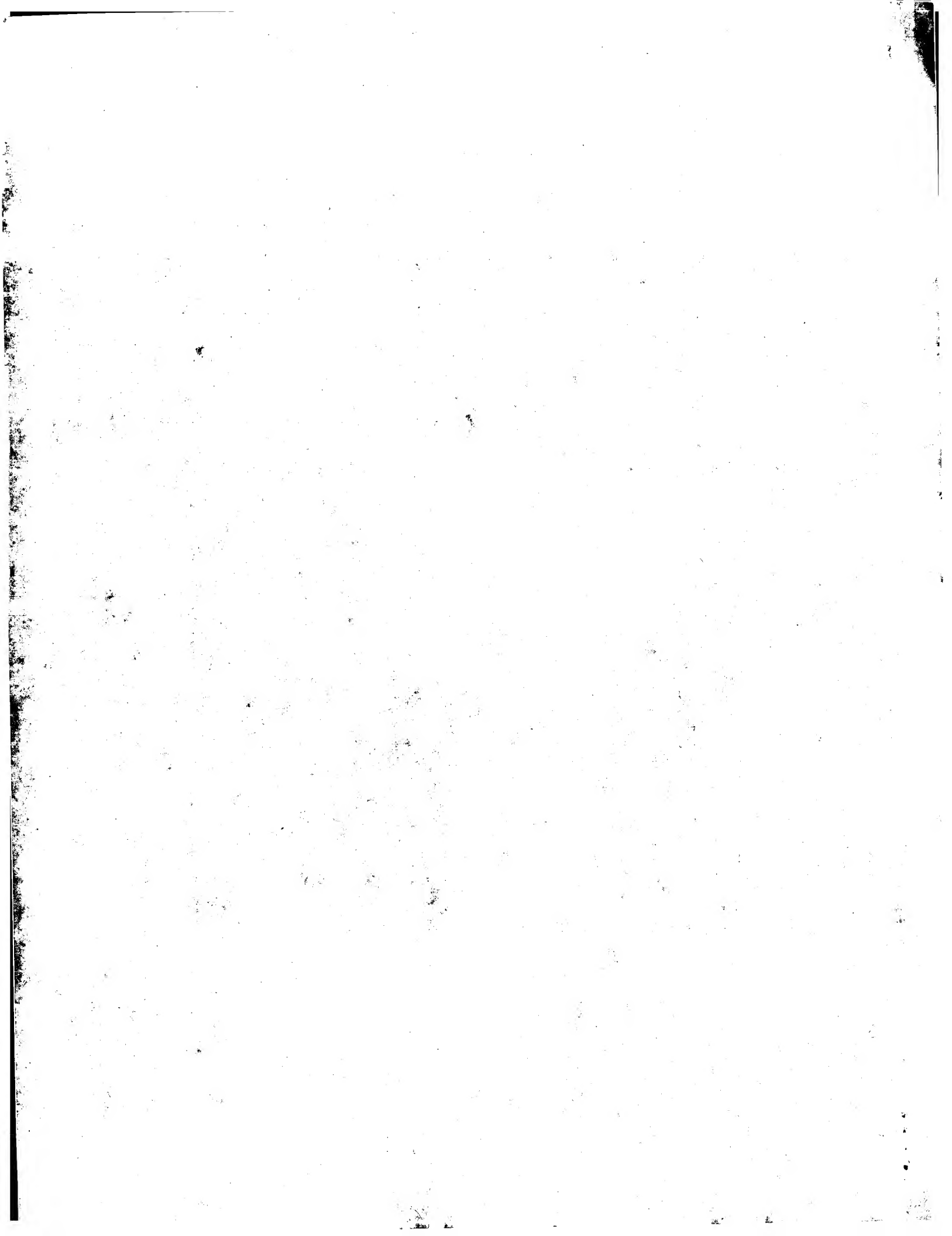


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AJ224333; CAB64688.1; -  
DR InterPro; IPR001813; 60s.Ribosomal.  
DR InterPro; IPR001859; Ribosomal\_P2.  
DR Pfam; PF00428; 60s.Ribosomal; 1.  
DR PRINTS; PR00456; RIBOSOMALP2.  
DR Ribosomal protein; Phosphorylation; Allergen.  
KW Ribosome 111 AA; 11136 MW; QFCDE3F6023994A7 CRC64;  
SQ SEQUENCE 111 AA; 11136 MW; QFCDE3F6023994A7 CRC64;  
  
Query Match 42.4%; Score 14; DB 1; Length 111;  
Best Local Similarity 17.6%; Pred. No. 3e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
QY 2 GGXXXXXXXG 18  
||  
DB 68 GGAAPAAAGAGG 84  
  
RESULT 14  
19KD\_MYCIT STANDARD; PRT; 162 AA.  
ID 19KD\_MYCIT  
AC F31502;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE 19 KDA LIPOPROTEIN ANTIGEN PRECURSOR (22 KDA LIPOPROTEIN ANTIGEN)  
DE (M122 ANTIGEN).  
GN M122.  
OS Mycobacterium intracellulare.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1767;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=DARDEN / SEROVAR 19;  
RC MEDLINE=93202760; PubMed=8454357;  
RX Booth R.J., Williams D.L., Moudgil K.D., Noonan L.C.,  
RA Grandison P.M., McKee J.J., Prestidge R.L., Watson J.D.;  
RT Homologs of Mycobacterium leprae 18-kilodalton and Mycobacterium  
RT tuberculosis 19-kilodalton antigens in other mycobacteria.";  
RL Infect. Immun. 61:1509-1515(1993).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=SEROVAR 14;  
RC MEDLINE=92326626; PubMed=1445568;  
RX Nair J., Rouse D.A., Morris S.L.;  
RA "Nucleotide sequence analysis and serologic characterization of the  
RT Mycobacterium intracellulare homologue of the Mycobacterium  
RT tuberculosis 19 kda antigen.";  
RL Mol. Microbiol. 6:1431-1439(1992).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID  
CC ANCHOR (PROBABLE).  
CC -1- SIMILARITY: TO OTHER MYCOBACTERIUM 19 KDA ANTIGEN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L12238; AAA25344.1; -  
DR EMBL; X65483; CAA46469.1; -  
DR PIR; S22630; S22630.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.

KW Antigen; Membrane; Lipoprotein; Signal.  
FT SIGNAL 1 21 PROBABLE.  
FT CHAIN 22 162 19 KDA LIPOPROTEIN ANTIGEN.  
FT LIPID 22 22 N-ACYL DIGLYCERIDE (PROBABLE).  
FT CONFLICT 37 37 N -> S (IN REF. 2).  
FT CONFLICT 42 48 TRRLAPG -> SASASTGG (IN REF. 2).  
SQ SEQUENCE 162 AA; 15517 MW; A38EC8100D8870C5 CRC64;  
  
Query Match 42.4%; Score 14; DB 1; Length 162;  
Best Local Similarity 17.6%; Pred. No. 4e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
QY 2 GGXXXXXXXG 18  
||  
DB 24 GGNKSGTSASSANSNG 40  
  
RESULT 15  
CH18\_DROME STANDARD; PRT; 172 AA.  
ID CH18\_DROME  
AC P07184;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE CHORION PROTEIN S18.  
DE CP18 OR S18.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=85229804; PubMed=2988878;  
RX Wong Y.-C., Pustell J., Spoerel N., Kafatos F.C.;  
RA "Coding and potential regulatory sequences of a cluster of chorion  
RT genes in Drosophila melanogaster.";  
RL Chromosoma 92:124-135(1985).  
[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=85229805; PubMed=3924529;  
RX Levine J., Spradling A.;  
RA "DNA sequence of a 3.8 kilobase pair region controlling Drosophila  
RT chorion gene amplification.";  
RL Chromosoma 92:136-142(1985).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X02497; CAA26328.1; -  
DR EMBL; X06257; CAA29602.1; -  
DR FlyBase; FBgn0000357; Cp18.  
DR Chorion.  
SQ SEQUENCE 172 AA; 17269 MW; 928405D3436D0DCE CRC64;  
  
Query Match 42.4%; Score 14; DB 1; Length 172;  
Best Local Similarity 17.6%; Pred. No. 4.2e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
  
QY 2 GGXXXXXXXG 18  
||  
DB 150 GGSAAASAAASVAAGKKG 166  
  
Search completed: February 12, 2002, 12:39:50

Job time: 803 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:37 ; Search time 232.64 Seconds  
(without alignments)  
11.317 Million cell updates/sec

Title: US-09-485-571-20

Perfect score: 33

Sequence: 1 XGGXXXXXXXXXXXXXG 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	54.5	286	13 Q9PUX6	Q9pux6 gadus morhu
2	18	54.5	526	5 Q9VXV2	Q9vxv2 drosophila
3	17	51.5	127	5 Q9VKR9	Q9vkr9 drosophila
4	17	51.5	186	5 Q9VIT6	Q9vit6 drosophila
5	17	51.5	357	10 Q9AWI0	Q9awi0 oryza sativ
6	17	51.5	513	11 Q9Z2W7	Q9z2w7 rattus norv
7	17	51.5	585	11 P70581	P70581 rattus norv
8	17	51.5	738	5 O02402	O02402 pinctada fu
9	17	51.5	803	4 Q15022	Q15022 homo sapien
10	16	48.5	121	5 Q9V5U8	Q9v5u8 drosophila
11	16	48.5	178	3 Q9URS5	Q9urs5 kluyveromyc
12	16	48.5	242	11 Q99LS8	Q99ls8 mus musculu
13	16	48.5	309	11 Q9CS23	Q9csz3 mus musculu
14	16	48.5	342	5 Q9VKR8	Q9vkr8 drosophila
15	16	48.5	347	13 Q9IB14	Q9ib14 gallus gall
16	16	48.5	361	2 Q9CI25	Q9ci25 lactococcus
17	16	48.5	430	13 Q9IB15	Q9ib15 gallus gall
18	16	48.5	476	12 Q80890	Q80890 herpesvirus
19	16	48.5	526	4 Q9UK58	Q9uk58 homo sapien

20	16	48.5	527	11 Q9RIQ2	Q9rlq2 rattus norv
21	16	48.5	602	10 Q9AR00	Q9ar00 lycopersico
22	16	48.5	605	5 Q9TYL3	Q9tyl3 caenorhabdi
23	16	48.5	623	10 Q9ST59	Q9st59 triticum ae
24	16	48.5	691	5 Q9BIU3	Q9biu3 dolomedes t
25	16	48.5	904	5 Q76271	Q76271 mytilus edu
26	16	48.5	904	5 Q9VC33	Q9vc33 drosophila
27	16	48.5	922	5 Q44367	Q44367 mytilus edu
28	16	48.5	1436	5 Q9VX21	Q9vx21 drosophila
29	16	48.5	1468	5 Q9GUB5	Q9gub5 gallieria me
30	15	45.5	88	5 Q9WLU9	Q9wlu9 drosophila
31	15	45.5	90	5 Q9NLI4	Q9nli4 leishmania
32	15	45.5	111	3 Q9HFQ4	Q9hfq4 candida alb
33	15	45.5	138	13 Q9XYW9	Q9xyw9 petromyzon
34	15	45.5	139	5 Q9W2W0	Q9w2w0 drosophila
35	15	45.5	169	4 Q9P0D7	Q9p0d7 homo sapien
36	15	45.5	251	4 Q9BXR5	Q9bxr5 homo sapien
37	15	45.5	251	11 Q9D4Z4	Q9d4z4 mus musculu
38	15	45.5	284	5 Q9BIS8	Q9bis8 tetragantha
39	15	45.5	310	5 Q9W496	Q9w496 drosophila
40	15	45.5	346	13 Q919B6	Q919b6 brachydanio
41	15	45.5	349	4 Q9BTW4	Q9btw4 homo sapien
42	15	45.5	360	5 Q9BIU0	Q9biu0 latrodectus
43	15	45.5	373	5 Q9BIT9	Q9bit9 latrodectus
44	15	45.5	377	10 Q9ZNY6	Q9zny6 oryza sativ
45	15	45.5	425	5 Q9BMT6	Q9bmt6 clona intes

#### ALIGNMENTS

RESULT 1  
Q9PUX6 ID Q9PUX6 PRELIMINARY; PRT; 286 AA.  
AC Q9PUX6  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE MIXED LINEAGE LEUKEMIA-LIKE PROTEIN (FRAGMENT).  
GN MLL.  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;  
OC Gadus.  
OX NCBI\_TaxID=8049;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99398697; PubMed=10468597;  
RA Venkatesh B., Ning Y., Brenner S.;  
RT "Late changes in spliceosomal introns define clades in vertebrate evolution."  
RT Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).  
RL PROC. NATL. ACAD. SCI. U.S.A. 96:10267-10271(1999).  
DR EMBL; AF137232; AAD53455.1;  
FT NON\_TER 1 1  
FT NON\_TER 286 286  
SQ SEQUENCE 286 AA; 29687 MW; BE724A520927BD8A CRC64;

Query Match: 54.5%; Score 18; DB 13; Length 286;  
Best Local Similarity 17.6%; Pred. No. 3.3e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 209 GGAAAAAAAAAAAAAG 225

RESULT 2  
Q9VXV2 ID Q9VXV2 PRELIMINARY; PRT; 526 AA.  
AC Q9VXV2  
DT 01-MAY-2000 (TReMBLrel. 13, Created)

01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 CG7294 PROTEIN.

DT DE GN  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Palazzolo M., Pittman K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003499; AAF48456.1;  
 DR FlyBase: FBgn0030640; CG6294.  
 DR InterPro: IPR000130; Zn.MTpeptidase.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN.1.  
 SQ SEQUENCE 526 AA; 58657 MW; 2DF254910BADF1F CRC64;

Query Match 54.5%; Score 18; DB 5; Length 526;  
 Best Local Similarity 17.6%; Pred.No. 5.9e+03;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXXXXXX 18  
 ||  
 DB 365 GGSSSSSSSSSSSSAG 381

RESULT 3  
 ID Q9VWR9 PRELIMINARY; PRT; 127 AA.  
 AC Q9VWR9;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 CG7294 PROTEIN.

DT DE GN  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Palazzolo M., Pittman K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003629; AAF52991.1;  
 DR FlyBase: FBgn0032284; CG7294.  
 DR InterPro: IPR000817; Prion.  
 DR PRINTS: PR00341; PRION.  
 SQ SEQUENCE 127 AA; 11014 MW; CBFDEC3362C2560F CRC64;

Query Match 51.5%; Score 17; DB 5; Length 127;  
 Best Local Similarity 17.6%; Pred.No. 2.6e+03;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXXXXXX 18  
 ||  
 DB 88 GGGAASASASSASAAG 104

RESULT 4  
 ID Q9VII6 PRELIMINARY; PRT; 186 AA.  
 AC Q9VII6;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)



```

DE CG14404 PROTEIN.
GN CG14404.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Abayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhargava P., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector E., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AF003668; AAF53933.1; -.
DR FlyBase: FBgn0032893; CG14404.
SQ SEQUENCE 186 AA; 21236 MW; D847BE4817AA7C79 CRC64;

```

```

Query Match 51.5%; Score 17; DB 5; Length 186;
Best Local Similarity 17.6%; Pred. No. 3.7e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 2 GGXXXXXXXXXXXXXG 18
   ||
DB 160 GGATTTTNTTTTTS 176

```

```

RESULT 5
Q9AW10 PRELIMINARY; PRT; 357 AA.
ID Q9AW10;
AC Q9AW10;

```

```

DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE PUTATIVE ZINC-FINGER PROTEIN.
GN P0489A05.20.
OS Oryza sativa (Rice).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone: P0489A05.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003105; BAB32997.1; -.
KW ZINC.
SQ SEQUENCE 357 AA; 37090 MW; 32F6C145E4823760 CRC64;

```

```

Query Match 51.5%; Score 17; DB 10; Length 357;
Best Local Similarity 17.6%; Pred. No. 7e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 2 GGXXXXXXXXXXXXXG 18
   ||
DB 261 GGRSSSSSSSSAAAG 277

```

```

RESULT 6
Q922W7 PRELIMINARY; PRT; 513 AA.
ID Q922W7;
AC Q922W7;

```

```

DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE P45.
GN P58/P45.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99013879; PubMed=9795236;
RA Hu T., Gerace L.;
RT "cDNA cloning and analysis of the expression of nucleoporin p45.";
RL Gene 221:245-253(1998).
DR EMBL: AF000900; AAC82318.1; -.
SQ SEQUENCE 513 AA; 51820 MW; C70EC7465F5AAD0B CRC64;

```

```

Query Match 51.5%; Score 17; DB 11; Length 513;
Best Local Similarity 17.6%; Pred. No. 1e+04;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 2 GGXXXXXXXXXXXXXG 18
   ||
DB 77 GGTSAAGTTATTSATTTG 93

```

```

RESULT 7
P70581 PRELIMINARY; PRT; 585 AA.
ID P70581;
AC P70581;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE NUCLEOPORIN P58.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96326666; PubMed=8707840;
RA Hu T., Guan T., Gerace L.;

```

RT "Molecular and functional characterization of the p62 complex, an  
 RT assembly of nuclear pore complex glycoproteins.";  
 RL J. Cell Biol. 134:589-601(1996).  
 DR EMBL: U63839; AAC52789.1; -;  
 SQ SEQUENCE 585 AA; 59264 MW; 4D5F5C2744A75C3C CRC64;

Query Match 51.5%; Score 17; DB 11; Length 585;  
 Best Local Similarity 17.6%; Pred. No. 1.1e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
 ||  
 Db 77 GGTAGTTATTGASTTG 93

RESULT 8  
 Q02402 PRELIMINARY; PRT; 738 AA.  
 ID C02402;  
 AC 002402;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE INSOLUBLE PROTEIN.  
 OS Pinctada fucata.  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidea;  
 OC Pterioidea; Pterioidea; Pinctada.  
 OX NCBI\_TaxID=50426;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97320490; PubMed=9177341;  
 RA Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,  
 RA Nakashima K., Takahashi T.;  
 RT "Structures of mollusc shell framework proteins.";  
 RL Nature 387:563-564(1997).  
 DR EMBL: D86074; BAA20466.1; -;  
 SQ SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;

Query Match 51.5%; Score 17; DB 5; Length 738;  
 Best Local Similarity 17.6%; Pred. No. 1.4e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
 ||  
 Db 280 GGSAAAAAAGG 296

RESULT 9  
 Q15022 PRELIMINARY; PRT; 803 AA.  
 ID Q15022;  
 AC Q15022;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE ORF (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127530; PubMed=8590280;  
 RA Nagase T., Seki N., Tanaka A., Ishikawa K., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. IV.  
 RT the coding sequences of 40 new genes (K1AA0121-K1AA0160) deduced by  
 RT analysis of cDNA clones from human cell line KG-1.";  
 RL DNA Res. 2:167-174(1995).  
 DR EMBL: D63881; BAA09931.1; -;  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.  
 DR SMART: SM00355; Znf\_C2H2; 1.  
 FT NON\_TER 1

SQ SEQUENCE 803 AA; 89963 MW; CDFB901A35F29A7C CRC64;

Query Match 51.5%; Score 17; DB 4; Length 803;  
 Best Local Similarity 17.8%; Pred. No. 1.6e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
 ||  
 Db 113 GGSYASSSSSAAAAAG 129

RESULT 10  
 Q9V508 PRELIMINARY; PRT; 121 AA.  
 ID Q9V508;  
 AC Q9V508;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE CG9080 PROTEIN.  
 GN CG9080.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Shandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003826; AAC58699.1;  
 DR Flybase: FBgn003593; CG9080.  
 SQ SEQUENCE 121 AA; 11690 MW; 8047A2803CDA3338 CRC64;

Query Match 48.5%; Score 16; DB 5; Length 121;  
Best Local Similarity 17.6%; Pred. No. 4.3e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 103 GGSASASASSSWG 119

## RESULT 11

Q9URS5 PRELIMINARY; PRT; 178 AA.

AC Q9URS5;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE CRN1 HOMOLOGUE (FRAGMENT).  
GN CRN1.  
OS Kluyveromyces lactis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
OX NCBI\_TaxID=28985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2359/152;  
RA Bao W.G., Huo K.K., Li Y.Y., Fukuhara H.;  
RT "Protein disulfide isomerase genes of Kluyveromyces lactis.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ243960; CAB51777.1;  
FT NON\_TER 1  
SQ SEQUENCE 178 AA; 18698 MW; C3D8EB079CE4661A CRC64;

Query Match 48.5%; Score 16; DB 3; Length 178;  
Best Local Similarity 17.6%; Pred. No. 6.2e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 106 GGAAGAAAAAGAAAG 122

## RESULT 12

Q99LS8 PRELIMINARY; PRT; 242 AA.

AC Q99LS8;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE SIMILAR TO PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC002244; AAH02244.1;  
SQ SEQUENCE 242 AA; 26995 MW; B8B062AA144ABE62 CRC64;

Query Match 48.5%; Score 16; DB 11; Length 242;  
Best Local Similarity 17.6%; Pred. No. 8.4e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 19 GGSAAAAAAGGGAG 35

## RESULT 13

Q9CS23

ID Q9CS23 PRELIMINARY; PRT; 309 AA.  
AC Q9CS23;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE 2610030E23RIK PROTEIN (FRAGMENT).  
GN 2610030E23RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl F., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL: AK011629; BAB27744.1;  
DR MGD: MGI:1922664; 2610030E23RIK.  
DR InterPro: IPR000553; Cyclin.  
DR SMART: SM00385; CYCLIN; 2.  
FT NON\_TER 309  
SQ SEQUENCE 309 AA; 34255 MW; CEE35068FC302E7A CRC64;

Query Match 48.5%; Score 16; DB 11; Length 309;  
Best Local Similarity 17.6%; Pred. No. 1.1e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
DB 29 GGSSSGTTTTTTTTTG 45

## RESULT 14

Q9VKR8 PRELIMINARY; PRT; 342 AA.

AC Q9VKR8;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE CG17108 PROTEIN.  
GN CG17108.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George F.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
 RA Fouts C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Stoen-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AF003629; AAF52992.1; -;  
 DR FlyBase: FBgn0032285; CG17108.  
 DR InterPro: IPR002952; Eggshell.  
 DR PRINTS: PRO1228; EGGSHLL.  
 SQ SEQUENCE 342 AA; 26813 MW; E98DD0407300F7AB CRC64;

Query Match 48.5%; Score 16; DB 5; Length 342;  
 Best Local Similarity 17.6%; Pred. No. 1.2e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
 ||  
 Db 319 GGSGASASASASAG 335

RESULT 15  
 Q9IB14  
 ID Q9IB14 PRELIMINARY; PRT; 347 AA.  
 AC Q9IB14;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE PBX1B.  
 GN PBX1B.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=REED WHITE LEGHORN; TISSUE=LIMB BUD;  
 RA Yamegishi A., Kuroiwa A.;  
 RT "chicken PBX1A mRNA, complete cds."  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER HOMEBOX DOMAINS.

DR EMBL: AB043620; BAA96136.1; -;  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR000047; HTH\_repressr.  
 DR Pfam: PF00046; homeobox; 1.  
 DR PRINTS: PR00031; HTHREPRESSR.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KW DNA-binding; Homeobox; Nuclear protein.  
 SQ SEQUENCE 347 AA; 38540 MW; F01F96FC41287854 CRC64;

Query Match 48.5%; Score 16; DB 13; Length 347;  
 Best Local Similarity 17.6%; Pred. No. 1.2e+04;  
 Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
 ||  
 Db 124 GGSAASASASASAG 140

Search completed: February 12, 2002, 12:38:38  
 Job time: 751 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:31 ; Search time 242.57 Seconds  
(without alignments)  
5.497 Million cell updates/sec

Title: US-09-485-571-21

Perfect score: 80

Sequence: 1 RAARLGYYXXRFGXRVR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	92.5	18	20 AAW99408	Protegrin derivati
2	54	67.5	18	20 AAW99406	Protegrin derivati
3	44	55.0	283	21 AAY49438	M. tuberculosis ar
4	42	52.5	18	20 AAW99403	Protegrin derivati
5	42	52.5	18	21 AAY93616	Peptide which may
6	40	50.0	16	16 AAR78768	Protegrin peptide
7	40	50.0	18	16 AAR78776	Protegrin peptide
8	40	50.0	917	20 AAY39927	T. roseum DNA poly
9	39	48.8	18	16 AAR78778	Protegrin peptide
10	39	48.8	18	20 AAW99412	Protegrin derivati
11	39	48.8	18	21 AAY93177	Protegrin-like pep

12	39	48.8	18	21 AAY93179	Protegrin-like pep
13	39	48.8	18	21 AAY93615	Peptide which may
14	39	48.8	184	21 AAB40568	Human ORFX ORF332
15	39	48.8	200	21 AAY68705	Amino acid sequenc
16	39	48.8	500	22 AAB65766	Cysteine protease
17	37	46.2	18	18 AAW36429	Antimicrobial prot
18	37	46.2	18	18 AAW18151	Cationic, antimicr
19	37	46.2	18	18 AAW18152	Cationic, antimicr
20	37	46.2	18	18 AAW09084	Cationic, antimicr
21	37	46.2	18	18 AAW09085	Cationic, antimicr
22	37	46.2	72	22 AAM15837	Peptide #2271 enco
23	37	46.2	72	22 AAM28347	Peptide #2384 enco
24	37	46.2	72	22 AAM03575	Peptide #2257 enco
25	37	46.2	462	22 AAB87490	B. thuringiensis 44
26	37	46.2	2502	21 AAB00117	M. tuberculosis po
27	36	45.0	32	22 AAM21679	Peptide #8113 enco
28	36	45.0	32	22 AAM37988	Peptide #12025 enc
29	36	45.0	72	21 AAB38040	Fragment of human
30	36	45.0	211	21 AAB37993	Human secreted pro
31	36	45.0	317	20 AAY29262	Amino acid sequenc
32	36	45.0	358	21 AAV71307	Human orphan G pro
33	36	45.0	358	21 AAB02841	Human G-protein co
34	36	45.0	358	22 AAG64296	Human GTP-binding
35	36	45.0	358	22 AAE04568	Human G-protein co
36	36	45.0	358	22 AAE02934	Human LTB4 recepto
37	36	45.0	358	22 AAB73096	Human leukotriene
38	36	45.0	365	21 AAY81693	Human Wnt-6 protei
39	36	45.0	365	22 AAB88439	Human membrane or
40	36	45.0	365	22 AAB49769	Anyloid-beta prote
41	36	45.0	389	20 AAY57456	Human fishboy G-pr
42	36	45.0	389	21 AAY94477	Human HG07 protein
43	36	45.0	389	22 AAE02933	Human LTB4 recepto
44	36	45.0	494	22 AAG81171	Mycobacterium tube
45	36	45.0	1464	19 AAW79294	An antigen from de

ALIGNMENTS

RESULT 1

AAW99408

ID AAW99408 standard; peptide; 18 AA.

XX AAW99408;

XX AC

XX 08-JUN-1999 (first entry)

XX Protegrin derivative peptide SM2189.

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

XX Synthetic.

XX OS

XX Key

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX Modified-site

XX WPI; 1999-190034/16.  
 XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells  
 XX  
 PS Claim 7; Page 28; 37pp; French.  
 XX  
 CC This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.  
 XX Sequence 18 AA;  
 SQ

Query Match 92.5%; Score 74; DB 20; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-06;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RAARLGYYRXXRFGXRVR 18  
 Db 1 raarlgyyrxxrfgxrvgr 18  
 |||||  
 |||||

RESULT 2  
 AAW99406  
 ID AAW99406 standard; peptide; 18 AA.  
 XX  
 AC AAW99406;  
 XX  
 DT 08-JUN-1999 (first entry)  
 XX  
 DE Protegrin derivative peptide SM2187.  
 XX  
 KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.  
 XX  
 OS Synthetic.  
 OS  
 XX WO9907728-A2.  
 PN  
 XX 18-FEB-1999.  
 PD  
 XX 06-AUG-1998; 98WO-FR01757.  
 PF  
 XX 12-AUG-1997; 97FR-0010297.  
 PR  
 XX (SYNT-) SYNT:EM SA.  
 PA  
 XX Calas B, Chavanieu A, Grassy G, Kaczorek M;  
 PI  
 XX WPI; 1999-190034/16.  
 DR  
 XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells  
 PT  
 XX Claim 7; Page 28; 37pp; French.  
 PS  
 XX This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active

CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.  
 XX  
 SQ Sequence 18 AA;

Query Match 67.5%; Score 54; DB 20; Length 18;  
 Best Local Similarity 61.1%; Pred. No. 0.0061;  
 Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 RAARLGYYRXXRFGXRVR 18  
 Db 1 rggrlayllrfairvgr 18  
 | | | | |  
 | | | | |

RESULT 3  
 AAY49438  
 ID AAY49438 standard; Protein; 283 AA.

XX  
 AC AAY49438;  
 XX  
 DT 17-MAR-2000 (first entry)  
 XX  
 DE M. tuberculosis arylamine N-acetyltransferase (NAT) protein.  
 XX  
 KW Arylamine N-acetyltransferase; NAT; mycobacterium; tuberculosis;  
 KW leprosy.  
 XX  
 OS Mycobacterium tuberculosis.  
 OS  
 XX WO9961625-A1.  
 PN  
 XX 02-DEC-1999.  
 PD  
 XX 27-MAY-1999; 99WO-GB01692.  
 PF  
 XX 27-MAY-1998; 98GB-0011407.  
 PR  
 XX (ISIS-) ISIS INNOVATION LTD.  
 PA  
 XX Sim E, Payton M, Sinclair J;  
 PI  
 XX WPI; 2000-072627/06.  
 DR  
 XX N-PSDB; RAZ46695.

Novel isolated polypeptides used to develop products for the detection  
 and treatment of mycobacterial infections e.g. tuberculosis, leprosy or  
 cattle disease -

Claim 1; Fig 1; 61pp; English.

The invention provides a novel arylamine N-acetyltransferase (NAT)  
 protein from M. tuberculosis, M. smegmatis and Salmonella typhimurium.  
 The methods and products can be used for the detection of mycobacterium  
 and for the treatment of mycobacterial infections. They can also be used  
 for the design of compounds for treating mycobacterial infections, e.g.  
 tuberculosis. Administration of ligands of mycobacterial arylamine NAT,  
 together with isoniazid may increase its effectiveness against other  
 pathogenic mycobacteria, such as M. leprae which causes leprosy of  
 M. bovis which infects cattle. The present sequence represents the  
 M. tuberculosis arylamine NAT protein.

Sequence 283 AA;

Query Match 55.0%; Score 44; DB 21; Length 283;  
 Best Local Similarity 64.3%; Pred. No. 4;  
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;



QY 3 ARLGYRXXRFGXRV 16  
 | | | | | | | |  
 Db 82 aelgyrvrrfaarv 95

## RESULT 4

AAW99403  
 ID AAW99403 standard; peptide; 18 AA.

XX AC AAW99403;

XX DT 08-JUN-1999 (first entry)

XX DE Protegrin derivative peptide SM1738.

XX KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX OS Synthetic.

XX PN WO9907728-A2.

XX PD 18-FEB-1999.

XX PF 06-AUG-1998; 98WO-FR01757.

XX PR 12-AUG-1997; 97FR-0010297.

XX PA (SYNT-) SYNT:EM SA.

XX PI Calas B, Chavanleu A, Grassy G, Kaczorek M;

XX DR WPI; 1999-190034/16.

XX PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX PS Claim 7; Page 28; 37pp; French.

XX CC This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX SQ Sequence 18 AA;

Query Match 52.5%; Score 42; DB 20; Length 18;

Best Local Similarity 50.0%; Pred. No. 0.71;

Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVGR 18

| | | | | | | |

Db 1 rgrlrsysrrrrfsvsgr 18

## RESULT 5

AA93616  
 ID AA93616 standard; peptide; 18 AA.

XX AC AA93616;

XX DT 25-SEP-2000 (first entry)

XX

DE Peptide which may be linked to anticancer agents.

XX Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 KW cancer.

XX OS Unidentified.

XX PN WO200032237-A1.

XX PD 08-JUN-2000.

XX PF 26-NOV-1999; 99WO-FR02939.

XX PR 30-NOV-1998; 98FR-0015073.

XX PA (SYNT-) SYNT:EM SA.

XX PI Tamsamani J, Kaczorek M, Colin De Verdiere A;

XX DR WPI; 2000-412166/35.

XX CC New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells  
 XX Disclosure; Page 8; 34pp; French.

XX CC The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.

XX SQ Sequence 18 AA;

Query Match 52.5%; Score 42; DB 21; Length 18;

Best Local Similarity 50.0%; Pred. No. 0.71;

Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVGR 18

| | | | | | | |

Db 1 rgrlrsysrrrrfsvsgr 18

## RESULT 6

AAR78768  
 ID AAR78768 standard; peptide; 16 AA.

XX AC AAR78768;

XX DT 08-OCT-1995 (first entry)

XX DE Protegrin peptide sequence.

XX KW protegrin; antibiotic; antimicrobial; antiviral; antibacterial;

XX KW antifungal;

XX OS Synthetic.

XX PN WO9503325-A.

XX PD 02-FEB-1995.

XX PF 20-JUL-1994; 94WO-US08305.

XX PR 20-JUL-1993; 93US-0093926.

PR 26-JUL-1993; 93US-0095769.  
 PR 13-JAN-1994; 94US-0182483.  
 PR 17-MAY-1994; 94US-0243879.  
 XX (REGC ) UNIV CALIFORNIA.  
 PA  
 XX Harwig SSL, Kokryakov VN, Lehrer RL;  
 PI  
 XX WPI; 1995-075188/10.  
 DR

XX Antibiotic peptide-based cpds. designated protegrin(s) - are  
 PT useful for treating and preventing viral and microbial infections  
 PT and as preservatives  
 XX

PS Disclosure; Page 19; 80pp; English.

XX New peptides are disclosed which are designated "protegrins". The  
 CC peptides are useful as antibacterial, antiviral and antifungal agents in  
 CC both animals and plants. The peptides are 16-18 amino acids in length  
 CC and are characterised by four invariant Cys residues at positions 6, 8,  
 CC 13 and 15 and either (1) by a characteristic pattern of basic and  
 CC hydrophobic amino acids and/or (2) being isolatable from animal (e.g.  
 CC porcine) leukocytes; or analogues of these peptides in which 1-4 of the  
 CC Cys residues is/are replaced by hydrophobic or small amino acids. The  
 CC peptides can be produced synthetically and some can be produced  
 CC recombinantly or can be isolated and purified from their native sources.  
 CC The peptides can be modified by N-acylation and/or C-terminal amidation  
 CC or esterification, and can be in linear or cystine-bridged form. D-Amino  
 CC acid residues can be present.  
 CC The present sequence is a specific example of the protegrin  
 CC analogues in which Cys(6, 8, 13) have been replaced by Gly.  
 XX

SQ Sequence 16 AA;

Query Match 50.0%; Score 40; DB 16; Length 16;  
 Best Local Similarity 61.5%; Pred. No. 1.4; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 5;

QY 1 RAARLGYRXXRFG 13  
 I I I I I I I I  
 Db 1 rgrlgygrrrfg 13

RESULT 7  
 AAR78776  
 ID AAR78776 standard; peptide: 18 AA.  
 XX  
 AC AAR78776;

DT 08-OCT-1995 (first entry)

XX Protegrin peptide sequence.

DE protegrin; antibiotic; antimicrobial; antiviral; antibacterial;  
 KW antifungal;

XX Synthetic.

OS WO9503325-A.

XX 02-FEB-1995.

XX 20-JUL-1994; 94WO-US08305.

XX 20-JUL-1993; 93US-0093926.

PR 26-JUL-1993; 93US-0095769.

PR 13-JAN-1994; 94US-0182483.

PR 17-MAY-1994; 94US-0243879.

XX (REGC ) UNIV CALIFORNIA.

XX Harwig SSL, Kokryakov VN, Lehrer RL;

XX WPI; 1995-075188/10.  
 DR  
 XX Antibiotic peptide-based cpds. designated protegrin(s) - are  
 PT useful for treating and preventing viral and microbial infections  
 PT and as preservatives  
 XX  
 PS Disclosure; Page 19; 80pp; English.

XX New peptides are disclosed which are designated "protegrins". The  
 CC peptides are useful as antibacterial, antiviral and antifungal agents in  
 CC both animals and plants. The peptides are 16-18 amino acids in length  
 CC and are characterised by four invariant Cys residues at positions 6, 8,  
 CC 13 and 15 and either (1) by a characteristic pattern of basic and  
 CC hydrophobic amino acids and/or (2) being isolatable from animal (e.g.  
 CC porcine) leukocytes; or analogues of these peptides in which 1-4 of the  
 CC Cys residues is/are replaced by hydrophobic or small amino acids. The  
 CC peptides can be produced synthetically and some can be produced  
 CC recombinantly or can be isolated and purified from their native sources.  
 CC The peptides can be modified by N-acylation and/or C-terminal amidation  
 CC or esterification, and can be in linear or cystine-bridged form. D-Amino  
 CC acid residues can be present.  
 CC The present sequence is a specific example of the protegrin  
 CC analogues in which Cys(6, 8, 13, 15) have all been replaced by Ser.

XX Sequence 18 AA;

Query Match 50.0%; Score 40; DB 16; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 1.6;  
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVR 18  
 I I I I I I I I I I  
 Db 1 ratrisfsrrrfsvsvgr 18

RESULT 8

AAY39927  
 ID AAY39927 standard; Protein; 917 AA.

XX AAY39927;

DT 13-DEC-1999 (first entry)

XX T. roseum DNA polymerase I protein sequence.

DE DNA polymerase I; thermostable enzyme; PCR enzyme;  
 KW 3' to 5' proof reading activity.

XX Thermomicrobium roseum.

OS US5962296-A.

XX 05-OCT-1999.

XX 24-FEB-1998; 98US-0028361.

XX 24-FEB-1998; 98US-0028361.

XX (NEWE ) NEW ENGLAND BIOLABS INC.

XX Ettwiller L, Xu S;

XX WPI; 1999-579612/49.

DR N-PSDB; AAZ27471.

XX Pure thermostable DNA polymerase providing high fidelity DNA  
 PT amplification -

XX Claim 1; Fig 1; 16pp; English.

XX This sequence represents the Thermomicrobium roseum (ATCC 27502)

CC thermostable DNA polymerase I of the invention. The polymerase is useful  
 CC in the polymerase chain reaction (PCR). The polymerase has 3' to 5' proof  
 CC reading activity and therefore provides high fidelity amplification.

XX  
 SQ Sequence 917 AA;

Query Match 50.0%; Score 40; DB 20; Length 917;  
 Best Local Similarity 60.0%; Pred. No. 56;  
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXR 15  
 ||| ||| ||| |||  
 Db 789 raaelgyvetlfgrr 803

## RESULT 9

AAR78778  
 ID AAR78778 standard; peptide; 18 AA.

XX  
 AC AAR78778;

XX  
 DT 08-OCT-1995 (first entry)

XX  
 DE Protegrin peptide sequence.

XX  
 KW protegrin; antibiotic; antimicrobial; antiviral; antibacterial;  
 KW antifungal;

XX  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers

FT Misc-difference 9 /note= "D-form residue"

FT Misc-difference 18 /note= "D-form residue"

FT

XX  
 PN W09503325-A.

XX  
 PD 02-FEB-1995.

XX  
 PF 20-JUL-1994; 94WO-US08305.

XX  
 PR 20-JUL-1993; 93US-0093926.

XX  
 PR 26-JUL-1993; 93US-0095769.

XX  
 PR 13-JAN-1994; 94US-0182483.

XX  
 PR 17-MAY-1994; 94US-0243879.

XX  
 PA (REGC ) UNIV CALIFORNIA.

XX  
 PI Harwig SSL, Kokryakov VN, Lehrer RL;

XX  
 DR WPI; 1995-075188/10.

XX  
 PT Antibiotic peptide-based cpds. designated protegrin(s) - are  
 PT useful for treating and preventing viral and microbial infections  
 PT and as preservatives

XX  
 PS Disclosure; Page 19; 80pp; English.

XX  
 CC New peptides are disclosed which are designated "protegrins". The  
 CC peptides are useful as antibacterial, antiviral and antifungal agents in  
 CC both animals and plants. The peptides are 16-18 amino acids in length  
 CC and are characterised by four invariant Cys residues at positions 6, 8,  
 CC 13 and 15 and either (1) by a characteristic pattern of basic and  
 CC hydrophobic amino acids and/or (2) being isolatable from animal (e.g.  
 CC porcine) leukocytes; or analogues of these peptides in which 1-4 of the  
 CC Cys residues is/are replaced by hydrophobic or small amino acids. The  
 CC peptides can be produced synthetically and some can be produced  
 CC recombinantly or can be isolated and purified from their native sources.  
 CC The peptides can be modified by N-acylation and/or C-terminal amidation  
 CC or esterification, and can be in linear or cystine-bridged form. D-Amino  
 CC acid residues can be present.

CC The present sequence is a specific example of the protegrin  
 CC analogues in which Cys(6, 13) have been replaced by Val and Gly.

XX  
 SQ Sequence 18 AA;

Query Match 48.8%; Score 39; DB 16; Length 18;  
 Best Local Similarity 50.0%; Pred. No. 2.3;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXR 18  
 ||| : ||| |||  
 Db 1 ratrivicrrrrfgvcvgr 18

## RESULT 10

AAW99412  
 ID AAW99412 standard; peptide; 18 AA.

XX  
 AC AAW99412;

XX  
 DT 08-JUN-1999 (first entry)

XX  
 DE Protegrin derivative peptide SM2196.

XX  
 KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX  
 OS Synthetic.

XX  
 PN W09907738-A2.

XX  
 PD 18-FEB-1999.

XX  
 PF 06-AUG-1998; 98WO-FR01757.

XX  
 PR 12-AUG-1997; 97FR-0010297.

XX  
 PA (SYNT-) SYNT:EM SA.

XX  
 PI Calas B, Chavanieu A, Grassy G, Kaczorek M;

XX  
 DR WPI; 1989-190034/16.

XX  
 PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX  
 PS Claim 7; Page 28; 37pp; French.

XX  
 CC This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX  
 SQ Sequence 18 AA;

Query Match 48.8%; Score 39; DB 20; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXR 18  
 | | | | | | |

Db 1 rgrrlsyrtrfststgr 18

RESULT 11  
 AAY93177  
 ID AAY93177 standard; peptide; 18 AA.  
 AC AAY93177;  
 XX  
 XX 06-DEC-2000 (first entry)  
 XX  
 XX Protegrin-like peptide antibiotic Doxo-SynB1.  
 XX  
 XX Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
 XX blood-brain barrier; diagnostic; central nervous system; protegrin;  
 KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
 KW cancer; Parkinson's disease; depression; pain; meningitis.  
 XX  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Modified-site 1  
 FT /note= "linked to doxorubicin via a succinate  
 FT (-CO-(CH2)2-CO-) linker; optionally linked  
 FT to benzylpenicillin by a glycoamide linker"  
 XX  
 XX WO200032236-A1.  
 XX  
 XX 08-JUN-2000.  
 XX  
 XX 26-NOV-1999; 99WO-FR02938.  
 XX  
 XX 30-NOV-1998; 98FR-0015074.  
 XX  
 XX (SYNT-) SYNT:EM SA.  
 XX  
 XX Clair P, Kaczorek M, Tamsamani J;  
 XX  
 XX WPI; 2000-422871/36.  
 XX  
 XX Use of linear peptides as vectors for active ingredients, useful for  
 PT diagnosis and treatment of central nervous system diseases, can  
 PT transport agents passively across the blood-brain barrier  
 XX  
 XX Example I; Page 13; 54pp; French.  
 XX  
 XX The invention relates to the use of linear peptides, coupled to an active  
 CC agent, to prepare a composition able to cross the blood-brain barrier  
 CC for diagnosis or treatment of disorders localised in the central nervous  
 CC system. The linear peptide preferably has the formula: (a) X1- X16;  
 CC (b); BXXXXXXXBXXXXXB; or (c) BXXXXXXXBXXXXXB, where: each of X1-X16  
 CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
 CC be Trp; each B is aa containing a side chain that includes a basic group;  
 CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
 CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
 CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
 CC Peptides able to cross the BBB include protegrins, Antennapedia,  
 CC tachyplesins, transpontan, etc. Of these several families have cytolytic  
 CC effects and are termed peptide antibiotics. They fall into 3 main  
 CC categories based on their structure: (i) peptides with alpha-helices,  
 CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
 CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
 CC fall into the peptide antibiotic categories defined above: (a)-peptides  
 CC are based on the Antennapedia family peptides; (b)-peptides are based on  
 CC protegrins; and (c)-peptides are based on tachyplesins. This sequence  
 CC represents a synthetic linear peptide designed on peptides able to cross  
 CC the BBB and is conjugated to a doxorubicin molecule by a succinate  
 CC linker. The peptide may also be linked to a benzylpenicillin molecule  
 CC by a glycoamide linker.  
 CC Conjugates of the linear peptides and the active agent are particularly  
 CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
 CC Parkinson's diseases, depression, pain and meningitis, but also for

CC studying drug behaviour in BBB models.

XX Sequence 18 AA;

Query Match 48.8%; Score 39; DB 21; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVR 18

Db 1 rgrlsysrrfststgr 18

RESULT 12

AAY93179

ID AAY93179 standard; peptide; 18 AA.

AC AAY93179;

XX 06-DEC-2000 (first entry)

XX Protegrin-like peptide antibiotic Dal-SynB1.  
 XX

XX Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
 KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
 KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
 KW cancer; Parkinson's disease; depression; pain; meningitis; dalargin.  
 XX

OS Synthetic.

XX Key Location/Qualifiers

FH Cross-links 1

FT /note= "cross-links to a molecule of dalargin via  
 FT a disulphide linker"

XX WO200032236-A1.

XX 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02938.

XX 30-NOV-1998; 98FR-0015074.

XX (SYNT-) SYNT:EM SA.

XX Clair P, Kaczorek M, Tamsamani J;

XX WPI; 2000-422871/36.

XX Use of linear peptides as vectors for active ingredients, useful for  
 PT diagnosis and treatment of central nervous system diseases, can  
 PT transport agents passively across the blood-brain barrier

XX Example II; Page 20; 54pp; French.

XX The invention relates to the use of linear peptides, coupled to an active  
 CC agent, to prepare a composition able to cross the blood-brain barrier  
 CC for diagnosis or treatment of disorders localised in the central nervous  
 CC system. The linear peptide preferably has the formula: (a) X1- X16;  
 CC (b); BXXXXXXXBXXXXXB; or (c) BXXXXXXXBXXXXXB, where: each of X1-X16  
 CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
 CC be Trp; each B is aa containing a side chain that includes a basic group;  
 CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
 CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
 CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
 CC Peptides able to cross the BBB include protegrins, Antennapedia,  
 CC tachyplesins, transpontan, etc. Of these several families have cytolytic  
 CC effects and are termed peptide antibiotics. They fall into 3 main  
 CC categories based on their structure: (i) peptides with alpha-helices,  
 CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
 CC beta-sheets, e.g. protegrin, tachyplesins, defensins; (iii) peptides  
 CC with no major structure but containing bends due to the presence of  
 CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
 CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
 CC Peptides able to cross the BBB include protegrins, Antennapedia,  
 CC tachyplesins, transpontan, etc. Of these several families have cytolytic  
 CC effects and are termed peptide antibiotics. They fall into 3 main  
 CC categories based on their structure: (i) peptides with alpha-helices,  
 CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
 CC beta-sheets, e.g. protegrin, tachyplesins, defensins; (iii) peptides  
 CC with no major structure but containing bends due to the presence of

CC Pro residues, e.g. bactericins and PR39. The peptides of the invention  
 CC fall into the peptide antibiotic categories defined above: (a)-peptides  
 CC are based on the Antennapedia family peptides; (b)-peptides are based on  
 CC protegrins; and (c)-peptides are based on tachyplesins. This sequence  
 CC represents a synthetic linear peptide designed on peptides able to cross  
 CC the BBB and is conjugated to a daltargin molecule by a disulphide linker.  
 CC Conjugates of the linear peptides and the active agent are particularly  
 CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
 CC Parkinson's diseases, depression, pain and meningitis, but also for  
 CC studying drug behaviour in BBB models.

XX Sequence 18 AA;

Query Match 48.8%; Score 39; DB 21; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVR 18  
 Db 1 rggrlsysrrfrststgr 18  
 | | | | | | | |

RESULT 13  
 AAY93615  
 ID AAY93615 standard; peptide; 18 AA.

XX AAY93615;

XX 25-SEP-2000 (first entry)

DT Peptide which may be linked to anticancer agents.

DE Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 KW cancer.

KW Unidentified.

OS WO200032237-A1.

PN 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02939.

XX 30-NOV-1998; 98FR-0015073.

XX (SYNT-) SYNT:EM SA.

XX Tamsamaji J, Kaczorek M, Colin De Verdiere A;

XX WPI; 2000-412166/35.

XX New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells

XX Disclosure; Page 8; 34pp; French.

XX The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.

XX Sequence 18 AA;

Query Match 48.8%; Score 39; DB 21; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVR 18  
 Db 1 rggrlsysrrfrststgr 18  
 | | | | | | | |

RESULT 14

AAB40568  
 ID AAB40568 standard; Protein; 184 AA.

XX AAB40568;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF332 polypeptide sequence SEQ ID NO:664.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC74777.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 774; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:23 ; Search time 106.12 Seconds  
(without alignments)  
3.817 Million cell updates/sec

Title: US-09-485-571-21

Perfect score: 80  
Sequence: 1 RAARLGYRXXRFGXRVRG 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	40	50.0	16	1 US-08-182-483A-20	Sequence 20, Appl
2	40	50.0	16	1 US-08-243-879A-19	Sequence 19, Appl
3	40	50.0	16	1 US-08-499-523-40	Sequence 40, Appl
4	40	50.0	16	4 US-09-128-345-40	Sequence 40, Appl
5	40	50.0	18	1 US-08-182-483A-28	Sequence 28, Appl
6	40	50.0	18	1 US-08-243-879A-27	Sequence 27, Appl
7	40	50.0	18	1 US-08-499-523-48	Sequence 48, Appl
8	40	50.0	18	1 US-08-499-523-63	Sequence 63, Appl
9	40	50.0	18	1 US-08-499-523-67	Sequence 67, Appl
10	40	50.0	18	4 US-09-128-345-48	Sequence 48, Appl
11	40	50.0	18	4 US-09-128-345-63	Sequence 63, Appl
12	40	50.0	18	4 US-09-128-345-67	Sequence 67, Appl
13	39	48.8	18	1 US-08-182-483A-30	Sequence 30, Appl
14	39	48.8	18	1 US-08-499-523-50	Sequence 50, Appl
15	39	48.8	18	4 US-09-128-345-50	Sequence 50, Appl
16	37	46.2	18	2 US-08-752-852A-230	Sequence 230, Appl
17	36	45.0	18	1 US-08-499-523-53	Sequence 53, Appl
18	36	45.0	18	1 US-08-499-523-54	Sequence 54, Appl
19	36	45.0	18	1 US-08-499-523-58	Sequence 58, Appl
20	36	45.0	18	1 US-08-499-523-59	Sequence 59, Appl
21	36	45.0	18	4 US-09-128-345-53	Sequence 53, Appl
22	36	45.0	18	4 US-09-128-345-54	Sequence 54, Appl
23	36	45.0	18	4 US-09-128-345-58	Sequence 58, Appl
24	36	45.0	18	4 US-09-128-345-59	Sequence 59, Appl
25	35	43.8	18	1 US-08-499-523-57	Sequence 57, Appl
26	35	43.8	18	1 US-08-499-523-62	Sequence 62, Appl
27	35	43.8	18	4 US-09-128-345-57	Sequence 57, Appl

28	35	43.8	18	4 US-09-128-345-62	Sequence 62, Appl
29	35	43.8	286	4 US-08-964-127-4	Sequence 4, Appl
30	35	43.8	520	4 US-08-964-127-2	Sequence 2, Appl
31	35	43.8	905	2 US-08-574-959A-9	Sequence 9, Appl
32	35	43.8	905	4 US-09-357-014-9	Sequence 9, Appl
33	35	43.8	1135	2 US-08-574-959A-7	Sequence 7, Appl
34	35	43.8	1135	4 US-09-357-014-7	Sequence 7, Appl
35	34	42.5	18	2 US-08-752-852A-86	Sequence 86, Appl
36	34	42.5	336	4 US-09-105-537-27	Sequence 27, Appl
37	34	42.5	349	4 US-09-343-011B-1	Sequence 1, Appl
38	34	42.5	3729	2 US-08-804-227C-4	Sequence 4, Appl
39	34	42.5	11877	4 US-09-105-537-6	Sequence 6, Appl
40	33	41.2	16	1 US-08-182-483A-18	Sequence 18, Appl
41	33	41.2	16	1 US-08-243-879A-17	Sequence 17, Appl
42	33	41.2	16	1 US-08-499-523-38	Sequence 38, Appl
43	33	41.2	16	4 US-09-128-345-38	Sequence 38, Appl
44	33	41.2	18	1 US-08-499-523-65	Sequence 65, Appl
45	33	41.2	18	3 US-08-984-294-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1  
US-08-182-483A-20  
; Sequence 20, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KORZYAKOV, VLADIMIR N.  
; APPLICANT: HARMIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,483A  
; FILING DATE: 13-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-182-483A-20  
  
Query Match 50.0%; Score 40; DB 1; Length 16;  
Best Local Similarity 61.5%; Pred. No. 0.17;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 RAARLGYRXXRFG 13  
| | | | |  
DB 1 RGRRLGYRRRFG 13

RESULT 2  
 US-08-243-879A-19  
 ; Sequence 19, Application US/08243879A  
 ; Patent No. 5708145  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; TITLE OF INVENTION: A NEW PROTEGRIN  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington, DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/243.879A  
 ; FILING DATE: 17-MAY-1994  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.22  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030  
 ; INFORMATION FOR SEQ ID NO: 19:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 16 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-243-879A-19

Query Match 50.0%; Score 40; DB 1; Length 16;  
 Best Local Similarity 61.5%; Pred. No. 0.17; 5; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 5;

QY 1 RAARLGYRXXRFG 13  
 | | | | |  
 DB 1 RGRRLGYGRRRFG 13

RESULT 3  
 US-08-499-523-40  
 ; Sequence 40, Application US/08499523  
 ; Patent No. 5804558  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,523  
 ; FILING DATE: 07-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.24  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030  
 ; INFORMATION FOR SEQ ID NO: 40:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 16 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-499-523-40

Query Match 50.0%; Score 40; DB 1; Length 16;  
 Best Local Similarity 61.5%; Pred. No. 0.17; 5; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 5;

QY 1 RAARLGYRXXRFG 13  
 | | | | |  
 DB 1 RGRRLGYGRRRFG 13

RESULT 4  
 US-09-128-345-40  
 ; Sequence 40, Application US/09128345  
 ; Patent No. 6159936  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/128,345  
 ; FILING DATE: 03-AUG-1998  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura, A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 8067-0054-999  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 40:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 16 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-09-128-345-40

Query Match 50.0%; Score 40; DB 4; Length 16;  
Best Local Similarity 61.5%; Pred. No. 0.17;  
Matches 8; Conservative 0; Mismatches 5; Indels 0;

QY 1 RAARLGYRXXRFG 13  
| | | | |  
Db 1 RGRGLGYRRRFG 13

RESULT 5  
US-08-182-483A-28  
; Sequence 28, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,483A  
; FILING DATE: 13-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-182-483A-28

Query Match 50.0%; Score 40; DB 1; Length 18;  
Best Local Similarity 44.4%; Pred. No. 0.2;  
Matches 8; Conservative 2; Mismatches 8; Indels 0;

QY 1 RAARLGYRXXRFGXRVGR 18  
| | | : | | | |  
Db 1 RATRISFRRRFSVSVGR 18

RESULT 6  
US-08-243-879A-27  
; Sequence 27, Application US/08243879A  
; Patent No. 5708145  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: A NEW PROTEGRIN  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER

; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington, DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/243,879A  
; FILING DATE: 17-MAY-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-243-879A-27

Query Match 50.0%; Score 40; DB 1; Length 18;  
Best Local Similarity 44.4%; Pred. No. 0.2;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVGR 18  
| | | : | | | |  
Db 1 RATRISFRRRFSVSVGR 18

RESULT 7  
US-08-499-523-48  
; Sequence 48, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030

; INFORMATION FOR SEQ ID NO: 48:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 US-08-499-523-48

Query Match 50.0%; Score 40; DB 1; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 0.2;  
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVR 18  
 |||: ||||  
 Db 1 RATRISFRRRFSVSVGR 18

RESULT 8  
 US-08-499-523-63  
 ; Sequence 63, Application US/08499523  
 ; Patent No. 5804558  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KORYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,523  
 ; FILING DATE: 07-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.24  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030  
 ; INFORMATION FOR SEQ ID NO: 63:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: group(6, 8, 13, 15)  
 ; OTHER INFORMATION: /note= "X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"  
 US-08-499-523-63

Query Match 50.0%; Score 40; DB 1; Length 18;  
 Best Local Similarity 50.0%; Pred. No. 0.2;  
 Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVR 18  
 | ||| ||||  
 Db 1 RGGRLXYRRRRFXVXVGR 18

RESULT 9  
 US-08-499-523-67  
 ; Sequence 67, Application US/08499523  
 ; Patent No. 5804558  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KORYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,523  
 ; FILING DATE: 07-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.24  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030  
 ; INFORMATION FOR SEQ ID NO: 67:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: group(6, 8, 13, 15)  
 ; OTHER INFORMATION: /note= "X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"  
 US-08-499-523-67

Query Match 50.0%; Score 40; DB 1; Length 18;  
 Best Local Similarity 50.0%; Pred. No. 0.2;  
 Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVR 18  
 | ||| ||||  
 Db 1 RGGRLXYRRRRFXVXVGR 18

RESULT 10  
 US-09-128-345-48  
 ; Sequence 48, Application US/09128345  
 ; Patent No. 6159936  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KORYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York

STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-128-345-48

Query Match 50.0%; Score 40; DB 4; Length 18;  
Best Local Similarity 44.4%; Pred. No. 0.2;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRVRG 18  
Db 1 RATRISFRRRFSVSVGR 18

RESULT 11  
US-09-128-345-63  
Sequence 63, Application US/09128345  
Patent No. 6159936  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HAWIG, SYLVIA S.L.  
APPLICANT: KORRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 8, 13, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-63

Query Match 50.0%; Score 40; DB 4; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.2;  
Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRVRG 18  
Db 1 RGRRLXYRRRFXVXVGR 18

RESULT 12  
US-09-128-345-67  
Sequence 67, Application US/09128345  
Patent No. 6159936  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HAWIG, SYLVIA S.L.  
APPLICANT: KORRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 8, 13, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-67

Query Match 50.0%; Score 40; DB 4; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.2;  
Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 RAARLGYYXXRFGXRVGR 18  
| | | | |  
Db 1 RGGRLXXYRRRFXVAVGR 18  
  
RESULT 13  
US-08-182-483A-30  
; Sequence 30, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(9, 18)  
; OTHER INFORMATION: /note= "D form of amino acids"  
; US-08-499-523-50  
  
Query Match 48.8%; Score 39; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.3;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
  
QY 1 RAARLGYYXXRFGXRVGR 18  
| | | | |  
Db 1 RATRIVFCRRRFGVCVGR 18  
  
RESULT 15  
US-09-128-345-50  
; Sequence 50, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.

QY 1 RAARLGYYXXRFGXRVGR 18  
| | | | |  
Db 1 RGGRLXXYRRRFXVAVGR 18  
  
RESULT 13  
US-08-182-483A-30  
; Sequence 30, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,483A  
; FILING DATE: 13-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(9, 18)  
; OTHER INFORMATION: /note= "D form of amino acids"  
; US-08-182-483A-30  
  
Query Match 48.8%; Score 39; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.3;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
  
QY 1 RAARLGYYXXRFGXRVGR 18  
| | | | |  
Db 1 RATRIVFCRRRFGVCVGR 18  
  
RESULT 14  
US-08-499-523-50  
; Sequence 50, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:

```
;
;   REGISTRATION NUMBER: 30,742
;   REFERENCE/DOCKET NUMBER: 8067-0054-999
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-9741
;   TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 50:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 18 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: group(9, 18)
;   OTHER INFORMATION: /note= "D form of amino acids"
US-09-128-345-50
```

```
Query Match      48.8%; Score 39; DB 4; Length 18;
Best Local Similarity 50.0%; Pred. No. 0.3;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
```

```
Qy      1 RAARLGYRXXRFGXRVR 18
      |||: ||| |||
Db      1 RATRIVFCRRRFGVCVGR 18
```

```
Search completed: February 12, 2002, 12:32:23
Job time: 451 sec
```







Query Match 55.0%; Score 44; DB 2; Length 332;  
Best Local Similarity 64.3%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ARLGYRXXRFGXRV 16  
| | | | | | | | | |

Db 82 AELGYRVRFARV 95  
| | | | | | | | | |

RESULT 3

H69068  
cell surface glycoprotein (s-layer protein) related protein - Methanobacterium thermoautotrophicum  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999

C:Accession: H69068

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; K1, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: function  
A:Reference number: A69000; MUID:98037514

A:Accession: H69068

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1408 <MTH>

A:Cross-references: GB:AE000911; GB:AE000666; NID:g2622623; PIDN:AAH85988.1; PID:g262263

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1513

A:Start codon: TTG

Query Match 53.8%; Score 43; DB 2; Length 1408;  
Best Local Similarity 50.0%; Pred. No. 8.5;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ARLGYRXXRFGXRVGR 18  
| : | | | | : | : | |

Db 1319 AKLGYRTFKFTLKGR 1334

RESULT 4

E72403

aminomethyltransferase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: E72403

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing

A:Reference number: A72200; MUID:99287316

A:Accession: E72403

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <ARN>

A:Cross-references: GB:AE001706; GB:AE000512; NID:g4980707; PIDN:AAD35303.1; PID:g498070

C:Genetics:

A:Gene: TM0211

C:Superfamily: aminomethyltransferase

Query Match 51.2%; Score 41; DB 2; Length 364;  
Best Local Similarity 52.9%; Pred. No. 5.8;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRVG 17  
+ + | | | | | | | |

Db 290 RIARKGYEVLKNGRVG 306

RESULT 5

B82364

DNA polymerase I VC0108 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: B82364

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Seller, L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: B82364

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-934 <HEI>

A:Cross-references: GB:AE004101; GB:AE003852; NID:g9654497; PIDN:AAF93286.1; GSPDB:G

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0108

A:Map position: 1

C:Superfamily: DNA-directed DNA polymerase I

Query Match 51.2%; Score 41; DB 2; Length 934;  
Best Local Similarity 56.2%; Pred. No. 14;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRV 16  
| | | | | | | | | |

Db 813 RAAQLGYVETIFGRRL 828

RESULT 6

Q01HBC

hypothetical protein (gene N internal ORF) - bovine coronavirus (strain Mebus)

N:Alternate names: IORF protein

C:Species: bovine coronavirus

C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 16-Jul-1999

C:Accession: C26347

R:Lapps, W.; Hoque, B.G.; Brian, D.A.

Virology 157, 47-57, 1987

A:Title: Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein

A:Reference number: A94357; MUID:87151119

A:Accession: C26347

A:Molecule type: genomic RNA

A:Residues: 1-207 <LAP>

A:Cross-references: GB:M16620; NID:g323354; PIDN:AAA66398.1; PID:g807593

C:Superfamily: coronavirus gene N internal ORF

Query Match 50.0%; Score 40; DB 1; Length 207;  
Best Local Similarity 50.0%; Pred. No. 5.3;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRV 16  
| | | | | | | | | |

Db 150 RLRLGFLRLARYSLRV 165

RESULT 7

JQ1174

hypothetical protein (gene N internal ORF) - turkey coronavirus

N:Alternate names: IORF protein

C:Species: turkey coronavirus

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998

C:Accession: JQ1174

R:Verbeek, A.; Tijssen, P.

J. Gen. Virol. 72, 1659-1666, 1991

A:Title: Sequence analysis of the turkey enteric coronavirus nucleocapsid and membrane

A:Reference number: JQ1172; MUID:91311418

A:Accession: J01174  
A:Molecule type: genomic RNA  
A:Residues: 1-207 <VER>  
C:Experimental source: strain Minnesota  
C:Comment: The gene encoding this protein is located within the nucleocapsid (N) gene.  
C:Superfamily: coronavirus gene N internal ORF

Query Match 50.0%; Score 40; DB 1; Length 207;  
Best Local Similarity 50.0%; Pred. No. 5.3;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRV 16  
| | | | | | | | | |  
Db 150 RLRLGFLRLARYSLRV 165

RESULT 8  
S06869  
hypothetical protein (gene N internal ORF) - bovine coronavirus (strain F15)  
N:Alternate names: IORF protein  
C:Species: bovine coronavirus  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 16-Jul-1999  
C:Accession: S06869  
R:Cruciere, C.; Laporte, J.  
Ann. Inst. Pasteur Virol. 139, 123-138, 1988  
A:Title: Sequence and analysis of bovine enteritic coronavirus (F15) genome. I. - Sequen  
A:Reference number: S06399; MUID:89087718  
A:Accession: S06869  
A:Molecule type: genomic RNA  
A:Residues: 1-207 <CRU>  
A:Cross-references: EMBL:M36656; NID:g210700; PIDN:AAA42759.1; PID:g210702  
A:Note: The source is designated as bovine enteritic coronavirus  
C:Comment: The gene encoding this protein is located within the nucleocapsid (N) gene.  
C:Superfamily: coronavirus gene N internal ORF

Query Match 50.0%; Score 40; DB 1; Length 207;  
Best Local Similarity 50.0%; Pred. No. 5.3;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRV 16  
| | | | | | | | | |  
Db 150 RLRLGFLRLARYSLRV 165

RESULT 9  
A57261  
Probable transposase - Rhizobium meliloti insertion sequence ISRm5  
C:Species: Rhizobium meliloti  
C:Date: 19-Jan-1996 #sequence\_revision 19-Jan-1996 #text\_change 29-Sep-1999  
C:Accession: A57261  
R:Laberge, S.; Middleton, A.T.; Wheatcroft, R.  
J. Bacteriol. 177, 3133-3142, 1995  
A:Title: Characterization, nucleotide sequence, and conserved genomic locations of inser  
A:Reference number: A57261; MUID:95286494  
A:Accession: A57261  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-398 <LAB>  
A:Cross-references: GB:U08627; NID:g533460; PIDN:AAA81001.1; PID:g533461  
A:Experimental source: strain IZ450  
C:Genetics:  
A:Mobile element: insertion sequence ISRm5  
C:Superfamily: Mycobacterium tuberculosis probable IS1081 transposase  
C:Keywords: DNA binding

Query Match 50.0%; Score 40; DB 2; Length 398;  
Best Local Similarity 47.4%; Pred. No. 9.6;  
Matches 9; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 4 RLGYRXXRFG-----XRVGR 18

Db 55 RLGYRSGHYGRTLITRVGK 73  
| | | | | | | | | |  
RESULT 10  
F83561  
Probable DNA polymerase alpha chain PA0669 [imported] - Pseudomonas aeruginosa (str  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83561  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic  
A:Reference number: A82950; MUID:20437337  
A:Accession: F83561  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1031 <STO>  
A:Cross-references: GB:AE004502; GB:AE004091; NID:g9946547; PIDN:AAG04058.1; GSPDB  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0669

Query Match 50.0%; Score 40; DB 2; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RAARLGYR 8  
| | | | | | | |  
Db 35 RAARLGYR 42

RESULT 11  
A23172  
FMRFamide neuropeptide precursor - California sea hare  
C:Species: Aplysia californica (California sea hare)  
C:Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 21-Jul-2000  
C:Accession: A23172  
R:Schaefer, M.; Piccolotto, M.R.; Kreiner, T.; Kaldany, R.R.; Taussig, R.; Scheller,  
Cell 41, 457-467, 1985  
A:Title: Aplysia neurons express a gene encoding multiple FMRFamide neuropeptides.  
A:Reference number: A23172; MUID:85176970  
A:Accession: A23172  
A:Molecule type: mRNA  
A:Residues: 1-309 <SCH>  
A:Cross-references: GB:M11282; GB:M14957; NID:gl55751; PIDN:AAA27751.1; PID:gl55752  
C:Keywords: alternative splicing; neuropeptide

Query Match 48.8%; Score 39; DB 2; Length 309;  
Best Local Similarity 52.9%; Pred. No. 12;  
Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVG 17  
| | | | | | | | | |  
Db 265 RFMRFGKRFMRFGRSVG 281

RESULT 12  
A25790  
FMRFamide neuropeptide precursor - California sea hare  
C:Species: Aplysia californica (California sea hare)  
C:Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 31-Dec-1993  
C:Accession: A25790  
R:Taussig, R.; Scheller, R.H.  
DNA 5, 453-461, 1986  
A:Title: The Aplysia FMRFamide gene encodes sequences related to mammalian brain pe  
A:Reference number: A25790; MUID:87132918  
A:Accession: A25790  
A:Molecule type: DNA; mRNA

A:Residues: 1-597 <TAU>  
C:Genetics:  
A:Introns: 37/1  
C:Keywords: neuropeptide

Query Match 48.8%; Score 39; DB 2; Length 597;  
Best Local Similarity 52.9%; Pred. No. 21;  
Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
  
Qy 1 RAARLGYYXXRFGXRVG 17  
| | | | | | | | | |  
Db 553 RMRFGKRMREGRSVG 569

RESULT 13  
T03576  
conserved hypothetical protein - Rhodobacter capsulatus  
C:Species: Rhodobacter capsulatus  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999  
C:Accession: T03576  
R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.  
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997  
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SBI003  
A:Reference number: Z14955; MUID:97404404  
A:Accession: T03576  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-194 <VLC>  
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16229.1; PID:g3128377  
C:Genetics:  
A:Map position: 1  
C:Superfamily: conserved hypothetical protein MJ0201

Query Match 46.9%; Score 37.5; DB 2; Length 194;  
Best Local Similarity 47.4%; Pred. No. 14;  
Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;  
  
Qy 1 RAAR--LGYRXXRFGXRV 16  
| | | | | | | | | |  
Db 147 RSARPYGLGLLAREGARI 165

RESULT 14  
G83340  
hypothetical protein PA2440 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83340  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83340  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-320 <STO>  
A:Cross-references: GB:AE004671; GB:AE004091; NID:g9948483; PIDN:ARG05828.1; GSPDB:GN001  
C:Genetics:  
A:Gene: PA2440

Query Match 46.9%; Score 37.5; DB 2; Length 320;  
Best Local Similarity 63.2%; Pred. No. 22;  
Matches 12; Conservative 0; Mismatches 4; Indels 3; Gaps 2;  
  
Qy 1 RAA--RLGYRXXRFGXRVG 17  
| | | | | | | | | |  
Db 266 RAALGRGLGYRFA-FGARNG 283

RESULT 15  
S36519  
L2 protein - human papillomavirus type 34  
C:Species: human papillomavirus type 34  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
C:Accession: S36519  
R:Delius, H.; Hofmann, B.  
submitted to the EMBL Data Library, August 1993  
A:Description: Primer-directed sequencing of human papillomavirus types.  
A:Reference number: S36469  
A:Accession: S36519  
A:Molecule type: DNA  
A:Residues: 1-472 <DEL>  
A:Cross-references: EMBL:X74476; NID:g396989; PIDN:CAA52559.1; PID:g396995  
C:Superfamily: papillomavirus L2 protein  
C:Keywords: late protein

Query Match 46.9%; Score 37.5; DB 2; Length 472;  
Best Local Similarity 45.0%; Pred. No. 32;  
Matches 9; Conservative 2; Mismatches 6; Indels 3; Gaps 1;  
  
Qy 1 RAARLGYYXXRF---GXRVG 17  
| | | | | | | | | |  
Db 304 RVSRLGQRATMTTRSGKRIG 323

Search completed: February 12, 2002, 12:34:38  
Job time: 556 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:50 ; Search time 67.2 Seconds  
(without alignments)  
9.821 Million cell updates/sec

Title: US-09-485-571-21

Perfect score: 80

Sequence: 1 RAARLGYRXXRFGXRVR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	44	55.0	283	1 NAT_MYCTU	P56848 mycobacteri
2	41	51.2	364	1 GCST_THEMA	Q9wy34 thermotoga
3	40	50.0	207	1 YIOR_CVBF	P22654 bovine coro
4	40	50.0	207	1 YIOR_CVBF	P10525 bovine coro
5	40	50.0	207	1 YIOR_CVTFE	P26626 turkey ente
6	40	50.0	398	1 TRA5_RHME	Q52873 rhizobium m
7	39	48.8	309	1 FMRL_APLCA	P08020 aplisia cal
8	39	48.8	552	1 FMRL_APLCA	P08021 aplisia cal
9	38	47.5	495	1 MURC_MYCLE	P57994 mycobacteri
10	37.5	46.9	472	1 VL2_HPV34	P36758 human papil
11	37.5	46.9	524	1 VL2_HPV22	P50796 human papil
12	37.5	46.9	3023	1 POLG_TVNV	P09814 t genome po
13	37	46.2	395	1 NARK_BAGSU	P46907 bacillus su
14	37	46.2	552	1 GAA4_MOUSE	Q9d6f4 mus musculu
15	37	46.2	811	1 MUTS_THEAQ	Q56215 thermus aqu
16	37	46.2	817	1 MUTS_THEAQ	Q92ix6 thermus aqu
17	37	46.2	818	1 MUTS_THEAQ	Q56239 thermus aqu
18	36.5	45.6	519	1 VL2_HPV23	P50797 human papil
19	36	45.0	364	1 WNT6_MOUSE	P22727 mus musculu
20	36	45.0	364	1 WNT6_MOUSE	Q9y6f9 homo sapien
21	36	45.0	495	1 NAT_MYCTU	P56848 mycobacteri
22	36	45.0	1192	1 METH_MYCTU	Q33259 mycobacteri
23	36	45.0	1206	1 METH_MYCTU	Q49775 mycobacteri
24	35	43.8	74	1 MTRG_METKA	Q32868 methanopyru
25	35	43.8	298	1 TRGB_VIBCH	P25543 vibrio chol
26	35	43.8	413	1 COBL_PSEDE	P21921 pseudomonas
27	35	43.8	1129	1 AK11_RAT	Q62924 rattus norv
28	35	43.8	1876	1 PPSA_MYCTU	Q10977 mycobacteri
29	34.5	43.1	476	1 VL2_HPV32	P36757 human papil
30	34.5	43.1	477	1 VL2_HPV42	P27235 human papil
31	34	42.5	93	1 AAT_METEX	P52069 methylobact
32	34	42.5	194	1 GAR1_SCHPO	Q66975 schizosacch
33	34	42.5	282	1 MIND_CHLUV	P56346 chlorella v

#### ALIGNMENTS

##### RESULT 1

ID	NAT_MYCTU	STANDARD;	PRT;	283 AA.
AC	P56848;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	ARYLAMINE N-ACETYLTRANSFERASE (EC 2.3.1.5).			
GN	NAT OR RV3566C OR MT3671 OR MTCY06G11.13C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.F., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Stulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cdc 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai N.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	CHARACTERIZATION			
RX	MEDLINE=99139759; PubMed=9973365;			
RA	Payton M.A., Auty R., Delgado R.T., Everitt M., Sim E.;			
RT	"Cloning and characterization of arylamine N-acetyltransferase genes			
RT	from Mycobacterium smegmatis and Mycobacterium tuberculosis: increased			
RT	expression results in isoniazid resistance."			
RL	J. Bacteriol. 181:1343-1347(1999).			
CC	-!- FUNCTION: COULD HAVE A ROLE IN ACETYLYATING, AND HENCE			
CC	INACTIVATING, THE ANTITUBERCULAR DRUG ISONIAZID.			
CC	-!- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE = COA +			
CC	N-ACETYLARYLAMINE.			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC.			
CC	-!- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.			
CC	-!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A			
CC	FRAMESHIFT IN POSITION 104.			
CC	-----			

P19802 lymnaea sta  
P31571 escherichia  
Q60231 homo sapien  
Q9rh70 bradyrhizob  
P31009 drosophila  
Q80953 human papil  
P22155 rhesus papil  
Q9hk06 thermoplas  
Q9v318 drosophila  
Q44584 alcaligenes  
Q00742 homo sapien  
Q03513 synechococc

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; 292774; CAB07160.1; ALT\_FRAME.  
DR EMBL; AE007168; AAK48028.1; -.  
DR TIGR; MT3671; -.  
DR TubercuList; RV3566C; -.  
DR InterPro; IPR001447; Acetyltransf2.  
DR ProDom; PD002783; Acetyltransf2; 1.  
KW Transferase; Acyltransferase; Complete proteome.  
SQ SEQUENCE 283 AA; 31029 MW; 9C8D98E3256D088A CRC64;

Query Match 55.0%; Score 44; DB 1; Length 283;  
Best Local Similarity 64.3%; Pred. No. 0.41;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ARLGYRXXRFGXRV 16  
| | | | | | | |  
Db 82 AELGYRVRFAARV 95

RESULT 2  
GCST\_THEME  
ID GCST\_THEME STANDARD; PRT; 364 AA.  
AC Q9WY54;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROBABLE AMINOMETHYLTRANSFERASE (EC 2.1.2.10) (GLYCINE CLEAVAGE SYSTEM  
DE T PROTEIN).  
DE GCVT OR TM0211.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima."  
RL Nature 393:323-329(1999).

CC -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF  
CC GLYCINE (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: (6S)-TETRAHYDROFOLATE +  
CC S-AMINOMETHYLDIHYDROLIPOYLPROTEIN = (6R)-5,10-  
CC METHYLENETETRAHYDROFOLATE + NH(3) + DIHYDROLIPOYLPROTEIN.

CC -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:  
CC P, T, L, AND H (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE GCVT FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AE001706; AAD35303.1; -.  
DR TIGR; TM0211; -.  
DR InterPro; IPR002536; GCV\_T.

DR Pfam; PF01571; GCV\_T; 1.  
KW Transferase; AminoTransferase; Complete proteome.  
SQ SEQUENCE 364 AA; 40332 MW; 66F33CCC151CC6F1 CRC64;

Query Match 51.2%; Score 41; DB 1; Length 364;  
Best Local Similarity 52.9%; Pred. No. 1.9;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRV 17  
| | | | | | | |  
Db 290 RIARKGYEVLKNGERVG 306

RESULT 3  
YIOR\_CVBF  
ID YIOR\_CVBF STANDARD; PRT; 207 AA.  
AC P22654;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-DEC-1992 (Rel. 24, Last annotation update)  
DE HYPOTHETICAL PROTEIN IN NUCLEOCAPSID ORF (IORF).

OS Bovine coronavirus (strain F15).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11129;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89087718; PubMed=3207501;

RA Cruciere C., Laporte J.;  
RT "Sequence and analysis of bovine enteric coronavirus (F15) genome.  
RT I. Sequence of the gene coding for the nucleocapsid protein; analysis  
RT of the predicted protein."  
RL Ann. Inst. Pasteur Virol. 139:123-138(1988).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M36656; AAM42759.1; -.  
DR PIR; S06869; S06869.  
KW Hypothetical protein.  
SQ SEQUENCE 207 AA; 23001 MW; A4E5DE61171BAB50 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 207;  
Best Local Similarity 50.0%; Pred. No. 1.7;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRV 16  
| | | | | | | |  
Db 150 RLFLGFLRLARYSLRV 165

RESULT 4  
YIOR\_CVBM  
ID YIOR\_CVBM STANDARD; PRT; 207 AA.  
AC P10525;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-DEC-1992 (Rel. 24, Last annotation update)  
DE HYPOTHETICAL PROTEIN IN NUCLEOCAPSID ORF (IORF).

OS Bovine coronavirus (strain Mebus).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11132;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87151119; PubMed=3029965;



```

RA Lapps W.E., Hoque B.G., Brian D.A.;
RT "Sequence analysis of the bovine coronavirus nucleocapsid and matrix
RL protein genes.";
RL Virology 157:47-57(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16620; AAA66398.1; -.
DR PIR: C26347; Q01HBC.
KW Hypothetical protein.
SQ SEQUENCE 207 AA; 23054 MW; BE76DC4D663DD32A CRC64;

Query Match 50.0%; Score 40; DB 1; Length 207;
Best Local Similarity 50.0%; Pred. No. 1.7;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRV 16
| | | | | | | |
Db 150 RLRLGRLGRLARYSLRV 165

RESULT 5
YIOR_CVTKE STANDARD; PRT; 207 AA.
AC P26626;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE HYPOTHETICAL PROTEIN IN NUCLEOCAPSID ORF (IORF).
OS Turkey enteric coronavirus (TCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MINNESOTA;
RX MEDLINE=91311418; PubMed=1856695;
RA Verbeek A., Tijssen P.;
RT "Sequence analysis of the turkey enteric coronavirus nucleocapsid and
RT membrane protein genes: a close genomic relationship with bovine
RT coronavirus.";
RL J. Gen. Virol. 72:1659-1666(1991).
DR PIR: JQ1174; JQ1174.
KW Hypothetical protein.
SQ SEQUENCE 207 AA; 23005 MW; F69E2D8F2F006F77 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 207;
Best Local Similarity 50.0%; Pred. No. 1.7;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRV 16
| | | | | | | |
Db 150 RLRLGRLGRLARYSLRV 165

RESULT 6
TRA5_RHIME STANDARD; PRT; 398 AA.
AC Q52873;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT ISRM5.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

```

```

OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IZ450;
RX MEDLINE=95286494; PubMed=7768811;
RA Laberge S., Middleton A.T., Wheatcroft R.;
RT "Characterization, nucleotide sequence, and conserved genomic
RT locations of insertion sequence ISRM5 in Rhizobium meliloti.";
RL J. Bacteriol. 177:3133-3142(1995).
CC -1- FUNCTION: REQUIRED FOR THE TRANSPOSITION OF THE INSERTION ELEMENT.
CC -1- SIMILARITY: BELONGS TO THE MUTATOR FAMILY OF TRANSPOSASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U08627; AAA81001.1; -.
DR InterPro: IPR001207; Transpo_mutator.
DR Pfam: PF00872; Transpo_mutator; 1.
DR PROSITE: PS01007; TRANSPOSASE_MUTATOR; 1.
KW Transposable element; Transposition; DNA-binding; DNA recombination.
SQ SEQUENCE 398 AA; 45695 MW; 0A7AA811C82CFF51 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 398;
Best Local Similarity 47.4%; Pred. No. 3.1;
Matches 9; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 4 RLGYRXXRFG-----XRVGR 18
| | | | | | | |
Db 55 RLGYRSHYGRTLITRVGK 73

RESULT 7
FMRL_APLCA STANDARD; PRT; 309 AA.
ID FMRL_APLCA
AC P08020;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FMRFAMIDE NEUROPEPTIDE PRECURSOR (CLONE FMRF-1) (FRAGMENT).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea;
OC Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85176970; PubMed=3838698;
RA Schaefer M., Picciotto M.R., Kreiner T., Kaldany R.R., Taussig R.,
RA Scheller R.H.;
RT "Aplysia neurons express a gene encoding multiple FMRFamide
RT neuropeptides.";
RL Cell 41:457-467(1985).
CC -1- FUNCTION: FMRFAMIDE INDUCES CONTRACTIONS IN VISCERAL AND SOMATIC
CC MUSCULATURE AS WELL AS IN THE HEART.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT FMRFAMIDE PRECURSORS ARE
CC GENERATED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; M11282; AAA27751.1; -.
DR PIR; A23172; A23172.
DR InterPro; IPR002544; FARP.
DR Pfam; PF01581; FARP; 19.
KW Amidation; Cleavage on pair of basic residues; Repeat;
KW Neuropeptide; Alternative splicing.
FT NON_TER 1 1
FT PEPTIDE 7 10 FMRFAMIDE.
FT PEPTIDE 22 25 FMRFAMIDE.
FT PEPTIDE 38 41 FMRFAMIDE.
FT PEPTIDE 45 48 FMRFAMIDE.
FT PEPTIDE 61 64 FMRFAMIDE.
FT PEPTIDE 76 79 FMRFAMIDE.
FT PEPTIDE 91 94 FMRFAMIDE.
FT PEPTIDE 106 109 FMRFAMIDE.
FT PEPTIDE 122 125 FMRFAMIDE.
FT PEPTIDE 138 141 FMRFAMIDE.
FT PEPTIDE 154 157 FMRFAMIDE.
FT PEPTIDE 170 173 FMRFAMIDE.
FT PEPTIDE 186 189 FMRFAMIDE.
FT PEPTIDE 202 205 FMRFAMIDE.
FT PEPTIDE 218 221 FMRFAMIDE.
FT PEPTIDE 234 237 FMRFAMIDE.
FT PEPTIDE 250 253 FMRFAMIDE.
FT PEPTIDE 266 269 FMRFAMIDE.
FT PEPTIDE 273 276 FMRFAMIDE.
FT MOD_RES 10 10 AMIDATION (G-11 PROVIDE AMIDE GROUP).
FT MOD_RES 25 25 AMIDATION (G-26 PROVIDE AMIDE GROUP).
FT MOD_RES 41 41 AMIDATION (G-42 PROVIDE AMIDE GROUP).
FT MOD_RES 48 48 AMIDATION (G-49 PROVIDE AMIDE GROUP).
FT MOD_RES 64 64 AMIDATION (G-65 PROVIDE AMIDE GROUP).
FT MOD_RES 79 79 AMIDATION (G-80 PROVIDE AMIDE GROUP).
FT MOD_RES 94 94 AMIDATION (G-95 PROVIDE AMIDE GROUP).
FT MOD_RES 109 109 AMIDATION (G-110 PROVIDE AMIDE GROUP).
FT MOD_RES 125 125 AMIDATION (G-126 PROVIDE AMIDE GROUP).
FT MOD_RES 141 141 AMIDATION (G-142 PROVIDE AMIDE GROUP).
FT MOD_RES 157 157 AMIDATION (G-158 PROVIDE AMIDE GROUP).
FT MOD_RES 173 173 AMIDATION (G-174 PROVIDE AMIDE GROUP).
FT MOD_RES 189 189 AMIDATION (G-190 PROVIDE AMIDE GROUP).
FT MOD_RES 205 205 AMIDATION (G-206 PROVIDE AMIDE GROUP).
FT MOD_RES 221 221 AMIDATION (G-222 PROVIDE AMIDE GROUP).
FT MOD_RES 237 237 AMIDATION (G-238 PROVIDE AMIDE GROUP).
FT MOD_RES 253 253 AMIDATION (G-254 PROVIDE AMIDE GROUP).
FT MOD_RES 269 269 AMIDATION (G-270 PROVIDE AMIDE GROUP).
FT MOD_RES 276 276 AMIDATION (G-277 PROVIDE AMIDE GROUP).
SQ SEQUENCE 309 AA; 36213 MW; 6F268BD6C3F90908 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 309;
Best Local Similarity 52.9%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVG 17
Db 265 RFWRFGRPMRGRSVG 281

RESULT 8
FMR2_APLCA
ID FMR2_APLCA STANDARD; PRT; 552 AA.
AC P08021;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FMRFAMIDE NEUROPEPTIDE PRECURSOR (CLONE FMRF-2).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidia;
OC Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87132918; PubMed=3816495;
RA Taussig R., Scheller R.H.;

```

```

RT "The Aplysia FMRFamide gene encodes sequences related to mammalian
RT brain peptides.";
RL DNA 5:453-461(1986).
RN [2]
RP SEQUENCE OF 87-267 FROM N.A.
RX MEDLINE=85176970; PubMed=3838698;
RA Schaefer M., Picciotto M.R., Kreiner T., Kaldany R.-R., Taussig R.,
RA Scheller R.H.;
RA "Aplysia neurons express a gene encoding multiple FMRFamide
neuropeptides.";
RL Cell 41:457-467(1985).
CC -!- FUNCTION: FMRFAMIDE INDUCES CONTRACTIONS IN VISCERAL AND SOMATIC
CC MUSCULATURE AS WELL AS IN THE HEART.
CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT FMRFAMIDE PRECURSORS ARE
CC GENERATED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14958; AAA27755.1; -.
CC EMBL; M11283; AAB59169.1; -.
CC InterPro; IPR002544; FARP.
CC Pfam; PF01581; FARP; 29.
KW Amidation; Cleavage on pair of basic residues; Repeat;
KW Neuropeptide; Alternative splicing.
FT PEPTIDE 24 27 FMRFAMIDE.
FT PEPTIDE 108 111 FMRFAMIDE.
FT PEPTIDE 115 118 FMRFAMIDE.
FT PEPTIDE 153 156 FMRFAMIDE.
FT PEPTIDE 160 163 FMRFAMIDE.
FT PEPTIDE 175 178 FMRFAMIDE.
FT PEPTIDE 191 194 FMRFAMIDE.
FT PEPTIDE 198 201 FMRFAMIDE.
FT PEPTIDE 214 217 FMRFAMIDE.
FT PEPTIDE 229 232 FMRFAMIDE.
FT PEPTIDE 244 247 FMRFAMIDE.
FT PEPTIDE 259 262 FMRFAMIDE.
FT PEPTIDE 274 277 FMRFAMIDE.
FT PEPTIDE 289 292 FMRFAMIDE.
FT PEPTIDE 304 307 FMRFAMIDE.
FT PEPTIDE 319 322 FMRFAMIDE.
FT PEPTIDE 334 337 FMRFAMIDE.
FT PEPTIDE 349 352 FMRFAMIDE.
FT PEPTIDE 365 368 FMRFAMIDE.
FT PEPTIDE 381 384 FMRFAMIDE.
FT PEPTIDE 397 400 FMRFAMIDE.
FT PEPTIDE 413 416 FMRFAMIDE.
FT PEPTIDE 429 432 FMRFAMIDE.
FT PEPTIDE 445 448 FMRFAMIDE.
FT PEPTIDE 461 464 FMRFAMIDE.
FT PEPTIDE 477 480 FMRFAMIDE.
FT PEPTIDE 493 496 FMRFAMIDE.
FT PEPTIDE 509 512 FMRFAMIDE.
FT PEPTIDE 516 519 FMRFAMIDE.
FT MOD_RES 27 27 AMIDATION (G-28 PROVIDE AMIDE GROUP).
FT MOD_RES 111 111 AMIDATION (G-112 PROVIDE AMIDE GROUP).
FT MOD_RES 118 118 AMIDATION (G-119 PROVIDE AMIDE GROUP).
FT MOD_RES 156 156 AMIDATION (G-157 PROVIDE AMIDE GROUP).
FT MOD_RES 163 163 AMIDATION (G-164 PROVIDE AMIDE GROUP).
FT MOD_RES 178 178 AMIDATION (G-179 PROVIDE AMIDE GROUP).
FT MOD_RES 194 194 AMIDATION (G-195 PROVIDE AMIDE GROUP).
FT MOD_RES 201 201 AMIDATION (G-202 PROVIDE AMIDE GROUP).
FT MOD_RES 217 217 AMIDATION (G-218 PROVIDE AMIDE GROUP).
FT MOD_RES 232 232 AMIDATION (G-233 PROVIDE AMIDE GROUP).
FT MOD_RES 247 247 AMIDATION (G-248 PROVIDE AMIDE GROUP).
FT MOD_RES 262 262 AMIDATION (G-263 PROVIDE AMIDE GROUP).

```

```

DR EMBL: AL583920: CAC31296.1; -.
DR Leproma; ML0915; -.
DR InterPro: IPR000713; Mur_ligase.
DR Pfam: PF01225; Mur_ligase; 1.
DR Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
KW ATP-binding; Complete proteome.
ATP_BIND 128 128 ATP (POTENTIAL).
FT SEQUENCE 495 AA; 51588 MW; 46F816A650C847A7 CRC64;

Query Match 47.5%; Score 38; DB 1; Length 495;
Best Local Similarity 47.4%; Pred. No. 9;
Matches 9; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 RAARLGYRXXRFGX--RVG 17
   ||| | | | | | | |
DB 237 RTAELGIRVIRYGSDDRIG 255

RESULT 10
VL2_HPV34
ID VL2_HPV34 STANDARD; PRT; 472 AA.
AC P36758;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MINOR CAPSID PROTEIN L2.
DN GN L2.
OS Human papillomavirus type 34.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OC NCBI_TaxID=10613;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delius H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RC Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X74476; CAA52559.1; -.
CC PIR: S35519; S36519.
CC InterPro: IPR000784; Late_L2.
CC Pfam: P700513; late_protein_L2; 1.
CC Coat protein; Late protein.
CC SEQUENCE 472 AA; 50648 MW; 5118928245C63C70 CRC64;

Query Match 46.9%; Score 37.5; DB 1; Length 472;
Best Local Similarity 45.0%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 1 RAARLGYRXXRFX---GXRVG 17
   | | | | | | | | | |
DB 304 RVSRILGQRATMFTRSGRKIG 323

RESULT 11
VL2_HPV22
ID VL2_HPV22 STANDARD; PRT; 524 AA.
AC P50796;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MINOR CAPSID PROTEIN L2.
DN GN L2.

```

```
OS Human papillomavirus type 22.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=37954;
RN [1]
RP SEQUENCE FROM N.A.
RA Delius H.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31780; AAA79406.1; -
DR InterPro: IPR000784; Late_L2.
DR Pfam: PF00513; late_protein_L2; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 524 AA; 57619 MW; E9738FE2C3B19CDB CRC64;
-----
Query Match 46.9%; Score 37.5; DB 1; Length 524;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 6; Indels 3; Gaps 1;

QY 1 RAARLGYR--XXRFGXRVG 17
DB 339 RVSRLGQRASLSTRSGARVG 358
| ||| | | | | |
| ||| | | | | |

RESULT 12
POLG.TVMV STANDARD: PRT: 3023 AA.
AC P09814; Q84898; Q84899; Q84900; Q84901; Q84902;
DT 01-MAR-1989 (Rel. 10, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
DE 1 (SK1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (GK2);
DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.44) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP)].
OS Tobacco vein mottling virus (TMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Potyvirus.
OX NCBI_TaxID=12228;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86286553; PubMed=3737407;
RA Domier L.L., Franklin K.M., Shahabuddin M., Hellmann G.M.,
RA Overmeyer J.H., Hiramath S.T., Siaw M.F.E., Lomonossoff G.P.,
RA Shaw J.G., Rhoads R.E.;
RA "The nucleotide sequence of tobacco vein mottling virus RNA.";
RL Nucleic Acids Res. 14:5417-5430(1986).
RN [2]
RP REVISIONS.
RA Shaw J.G.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RX SEQUENCE OF 1874-1888, AND COVALENT RNA-LINKAGE.
RA MEDLINE=91087329; PubMed=1702164;
RA Murphy J.F., Rychlik W., Rhoads R.E., Hunt A.G., Shaw J.G.;
RA "A tyrosine residue in the small nuclear inclusion protein of tobacco
RT vein mottling virus links the VPg to the viral RNA.";
RL J. Virol. 65:511-513(1991).
CC -!- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
```

MAY BE INVOLVED IN REPLICATION.

-!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.

-!- CATALYTIC ACTIVITY: NUCLEAR INCLUSION PROTEIN A HYDROLYSES GLUTAMINYL BONDS, AND ACTIVITY IS FURTHER RESTRICTED BY PREFERENCES FOR THE AMINO ACIDS IN P6 -P1' THAT VARY WITH THE SPECIES OF POTYVIRUS, E.G. GLU-XAA-XAA-TYR-XAA-GLN+(SER OR GLY) FOR THE ENZYME FROM TOBACCO GLU-XAA-TYR-XAA-GLN+(SER OR GLY) THE VIRAL POLYPROTEIN, BUT OTHER PROTEINS AND OLIGOPEPTIDES CONTAINING THE APPROPRIATE CONSENSUS SEQUENCE ARE ALSO CLEAVED.

-!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.

-!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT INDIVIDUAL PROTEINS.

-!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.

-!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.

-!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

-----

EMBL: X04083; CAA27720.1; -

PIR: A23647; GNVSTV.

DR MEROPS: C04.001; -

DR MEROPS: C06.001; -

DR MEROPS: S30.001; -

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR001650; Helicase\_C.

DR InterPro: IPR001730; Peptidase\_C4.

DR InterPro: IPR001456; Peptidase\_C6.

DR InterPro: IPR002540; Poty\_P1.

DR InterPro: IPR001592; Poty\_coat.

DR InterPro: IPR001205; RNA\_pol\_P3D.

DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00270; DEAD; 1.

DR Pfam: PF00271; Helicase\_C; 1.

DR Pfam: PF00863; Peptidase\_C4; 1.

DR Pfam: PF00851; Peptidase\_C6; 1.

DR Pfam: PF00767; Poty\_coat; 1.

DR Pfam: PF01577; Poty\_P1; 1.

DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.

DR PRINTS: PR00966; NIAPOTYP\_TASE.

DR SMART: SM00487; DEXDC; 1.

DR SMART: SM00490; HELIC\_C; 1.

DR Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase; Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase; ATP-binding.

KW

KW

KW

FT CHAIN 1 265 N-TERMINAL PROTEIN.

FT CHAIN 266 749 HELPER COMPONENT PROTEINASE.

FT CHAIN 750 1130 29 KDA PROTEIN.

FT CHAIN 1131 1765 CYTOPLASMIC INCLUSION PROTEIN.

FT CHAIN 1766 1818 6 KDA PROTEIN.

FT CHAIN 1819 2242 NUCLEAR INCLUSION PROTEIN A.

FT CHAIN 2243 2758 NUCLEAR INCLUSION PROTEIN B.

FT CHAIN 2759 3023 COAT PROTEIN.

FT SITE 1130 1131 CLEAVAGE (BY 49 KDA PROTEASE).

FT SITE 1765 1766 CLEAVAGE (BY 49 KDA PROTEASE).

FT SITE 2242 2243 CLEAVAGE (BY 49 KDA PROTEASE).

FT SITE 2758 2759 CLEAVAGE (BY 49 KDA PROTEASE).

FT BINDING 1878 1878 COVALENT LINKAGE OF VIRAL RNA.

FT NP\_BIND 1215 1222 ATP (POTENTIAL).

SQ SEQUENCE 3023 AA; 342280 MW; 299FDED15C0E5B87 CRC64;

Query Match 46.9%; Score 37.5; DB 1; Length 3023;

Best Local Similarity 71.4%; Pred. No. 58;

Matches 10; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 5 LGYXXRFGXRVR 18  
 || | || || ||  
 Db 1476 LGERIQRFGRVGR 1488

RESULT 13  
 ID NARK\_BACSU STANDARD; PRT; 395 AA.  
 AC P46907;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NITRITE EXTRUSION PROTEIN (NITRITE FACILITATOR).  
 GN NARK.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96112813; PubMed=8846791;  
 RA Cruz Ramos H., Boursier L., Moszer I., Kunst F., Danchin A.,  
 Glaser P.;  
 RT "Anaerobic transcription activation in Bacillus subtilis:  
 RT Identification of distinct FNR-dependent and -independent regulatory  
 RT mechanisms.";  
 RL EMBO J. 14:5984-5994(1995).  
 RN [2]  
 RP SEQUENCE OF 1-171 FROM N.A.  
 RC STRAIN=168;  
 RA Presecan E., Moszer I., Boursier L., Cruz Ramos H.,  
 de la Fuente V., Hullo M., Lelong C., Schleich S., Sekowska A.,  
 Song B., Villani G., Kunst F., Danchin A., Glaser P.;  
 RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases  
 RL [1]  
 CC -1- FUNCTION: INVOLVED IN EXCRETION OF NITRITE PRODUCED BY THE  
 CC DISIMILATORY REDUCTION OF NITRATE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE NARK/NASA FAMILY OF TRANSPORTERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z49884; CAA90041.1; -;  
 DR EMBL; Z97024; CAB09704.1; -;  
 DR EMBL; Z99123; CAB15760.1; -;  
 DR Subtilist; BG11342; nark.  
 KW Transport; Transmembrane; Nitrate assimilation; Complete proteome.  
 FT TRANSMEM 15 35 POTENTIAL.  
 FT TRANSMEM 44 64 POTENTIAL.  
 FT TRANSMEM 73 93 POTENTIAL.  
 FT TRANSMEM 96 116 POTENTIAL.  
 FT TRANSMEM 133 153 POTENTIAL.  
 FT TRANSMEM 160 180 POTENTIAL.  
 FT TRANSMEM 203 223 POTENTIAL.  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT TRANSMEM 293 313 POTENTIAL.  
 FT TRANSMEM 330 350 POTENTIAL.  
 FT TRANSMEM 357 377 POTENTIAL.  
 SQ SEQUENCE 395 AA; 42956 MW; E41A35C3EC58921F CRC64;

Query Match 46.28; Score 37; DB 1; Length 395;  
 Best Local Similarity 58.3%; Pred. No. 11;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 LGYXXRFGXRVR 16  
 || | || || ||

Db 62 LGYLTNRFGARL 73

RESULT 14  
 ID GAA4\_MOUSE STANDARD; PRT; 552 AA.  
 AC Q9D6F4;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-4 SUBUNIT PRECURSOR (GABA(A)  
 DE RECEPTOR).  
 GN GABRA4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Hippocampus;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaído I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Guncinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE  
 CC VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE  
 CC GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE  
 CC CHANNEL.  
 CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)  
 CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; A013727; BAB28975.1; -;  
 DR MGD; MGI:95616; Gabra4.  
 DR InterPro; IPR001175; Neur\_channel.  
 DR Pfam; PF00065; neur\_chan; 1.  
 DR PRINTS; PR00252; NRIONCHANNEL.  
 DR PROSITE; PS00236; NEUOTR\_ION\_CHANNEL; 1.  
 DR Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;  
 KW Transmembrane; Multigene family.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 36 552 GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-4  
 FT SUBUNIT.  
 FT DOMAIN 36 258 EXTRACELLULAR (PROBABLE).  
 FT TRANSMEM 259 280 PROBABLE.  
 FT TRANSMEM 285 306 PROBABLE.  
 FT TRANSMEM 318 340 PROBABLE.

Search completed: February 12, 2002, 12:39:51  
Job time: 804 sec

```
FT DOMAIN 341 521 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 522 541 PROBABLE.
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 172 186 BY SIMILARITY.
SQ SEQUENCE 552 AA; 60878 MW; 213C16C423D7F97B CRC64;

Query Match 46.2%; Score 37; DB 1; Length 552;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 10; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

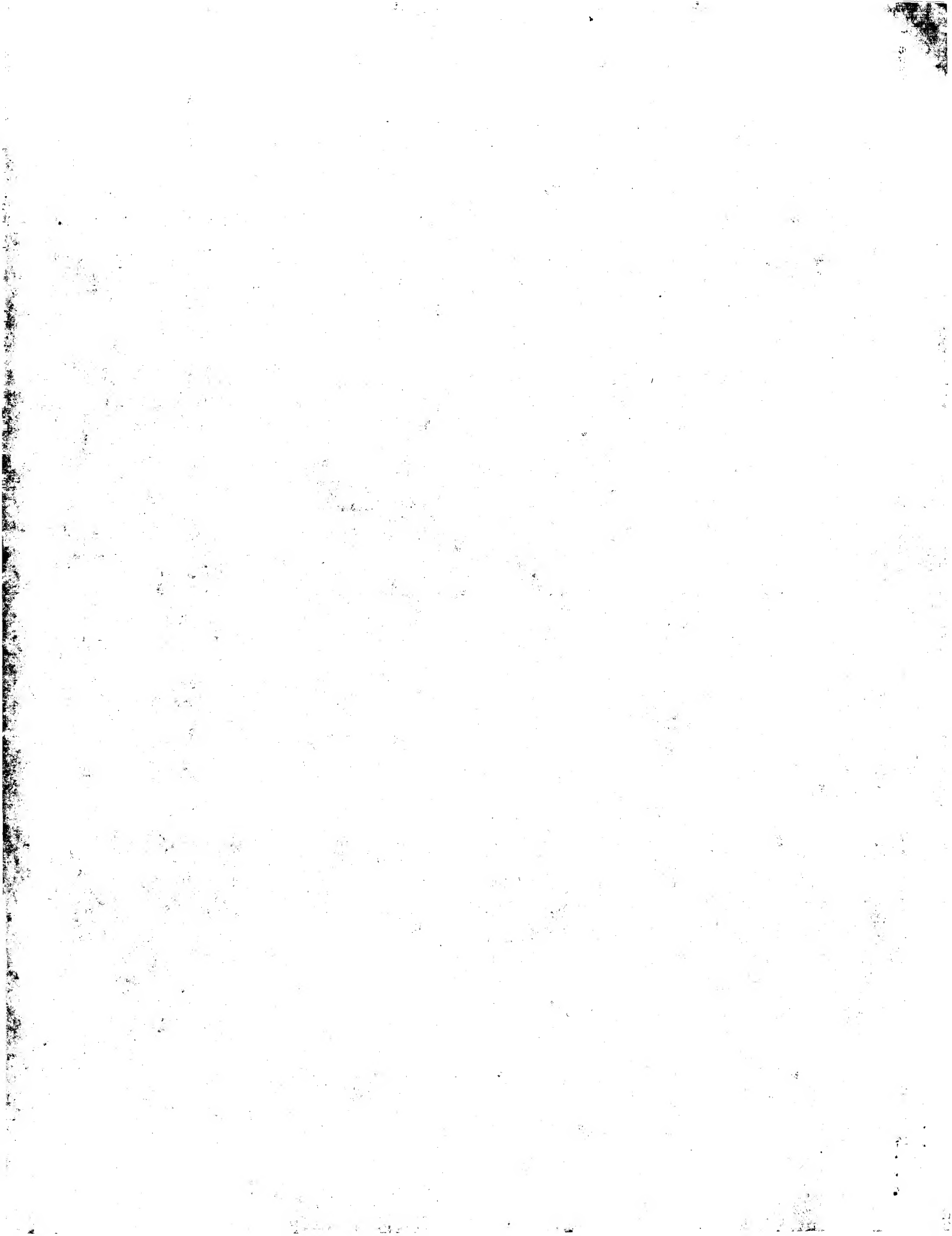
QY 1 RAARLGYRXXR--FCXRVGR 18
   | | | | | | | | | | | | | |
Db 464 RPASLGSASTRPAFGSRLGR 483

RESULT 15
MUTS_THEAQ
ID MUTS_THEAQ STANDARD; PRT; 811 AA.
AC Q56215;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA MISMATCH REPAIR PROTEIN MUTS.
GN MUTS.
OS Thermus aquaticus.
OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX NCBI_TaxID=271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YTL;
RX MEDLINE=96214934; PubMed=8617781;
RT Biswas I., Hsieh P.;
RT "Identification and characterization of a thermostable Muts homolog
   from Thermus aquaticus";
RL J. Biol. Chem. 271:5040-5048(1996).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES
CC IN DNA. IT IS POSSIBLE THAT IT CARRY OUT THE MISMATCH RECOGNITION
CC STEP. THIS PROTEIN HAS A WEAK ATPASE ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33117; AAC43637.1; -.
CC InterPro; IPR000432; Muts_C.
CC InterPro; IPR002863; Muts_N.
CC Pfam; PF00488; Muts_C; 1.
CC Pfam; PF01624; Muts_N; 1.
CC ProDom; PD001263; Muts_C; 1.
CC SMART; SM00534; Mutsac; 1.
CC SMART; SM00533; Mutsd; 1.
CC PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding.
FT NP_BIND 583 590 ATP (POTENTIAL).
SQ SEQUENCE 811 AA; 90627 MW; 43A5E5296A333515 CRC64;

Query Match 46.2%; Score 37; DB 1; Length 811;
Best Local Similarity 53.3%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 AARLGYRXXRFXRV 16
   | | | | | | | |
Db 537 AVRYGYVRPRFGDRL 551
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:38 ; Search time 232.64 Seconds  
(without alignments)  
11.317 Million cell updates/sec

Title: US-09-485-571-21  
Perfect score: 80  
Sequence: 1 RAARLGYXXRFGXRVR 18  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTEMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	45	56.2	433	5	O45311 caenorhabdi
2	43	53.8	492	5	O76267 aphysia cal
3	43	53.8	1408	1	O27557 methanobact
4	41	51.2	934	2	O9KVN3 vibrio chol
5	40	50.0	207	12	O9QAR7 Q9qar7 bovine coro
6	40	50.0	207	12	O9QAR0 Q9qar0 bovine coro
7	40	50.0	207	12	O9QAQ4 Q9qaq4 bovine coro
8	40	50.0	513	2	O9K4V0 alcaligenes
9	40	50.0	917	2	O9RHA4 Q9rha4 thermomicro
10	40	50.0	1031	2	O915O2 Q915g2 pseudomonas
11	39	48.8	127	5	O16963 aphysia cal
12	39	48.8	373	5	O16964 Q16964 aphysia cal
13	38	47.5	415	2	O68994 O68994 mycobacteri
14	38	47.5	421	2	O9CCS8 O9ccs8 mycobacteri
15	37.5	46.9	194	2	O68139 O68139 rhodobacter
16	37.5	46.9	320	2	O91142 O91142 pseudomonas
17	37.5	46.9	3023	12	O88925 Q88925 tobacco vei
18	37	46.2	224	1	P94950 P94950 methanopyru
19	37	46.2	253	5	O9N3G0 O9n3g0 caenorhabdi

20	37	46.2	256	1	O58492 pyrococcus
21	37	46.2	256	1	O9Y9K7 O9y9k7 aeropyrum p
22	37	46.2	289	4	O9H571 O9h571 homo sapien
23	37	46.2	396	13	O9W6S4 O9w6s4 gallus gall
24	37	46.2	417	2	O9R6S6 O9r6s6 synecococc
25	37	46.2	458	1	O9HQE5 O9hqes halobacteri
26	37	46.2	585	3	O9Y857 O9y857 kluyveromyc
27	37	46.2	644	2	O9RJK2 O9rjk2 streptomyc
28	37	46.2	1194	2	O9R6F0 O9r6f0 agrobacteri
29	36.5	45.6	321	12	O69545 O69545 human herpe
30	36	45.0	141	2	O66320 O66320 unidentified
31	36	45.0	243	2	O9JWN1 O9jwn1 neisseria m
32	36	45.0	279	2	O06057 O06057 mycobacteri
33	36	45.0	300	2	O49785 O49785 mycobacteri
34	36	45.0	334	2	O9XC63 O9xc63 pseudomonas
35	36	45.0	358	4	O9NPE5 O9npes homo sapien
36	36	45.0	359	3	P78762 P78762 schizosacch
37	36	45.0	359	8	O9TJR6 O9tjr6 prototheca
38	36	45.0	389	4	O9NPC1 O9npc1 homo sapien
39	36	45.0	391	1	O9HM39 O9hm39 thermoplasm
40	36	45.0	411	2	O06053 O06053 mycobacteri
41	36	45.0	414	2	O53038 O53038 rhodococcus
42	36	45.0	456	2	O9HWL6 O9hw16 pseudomonas
43	36	45.0	562	5	O9NKL3 O9nkl3 leishmania
44	36	45.0	718	10	O9SM79 O9sm79 oryza sativ
45	36	45.0	926	5	Q18284 Q18284 caenorhabdi

ALIGNMENTS

RESULT 1  
O45311 PRELIMINARY; PRT; 433 AA.  
AC O45311;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE C53B7.7 PROTEIN.  
GN C53B7.7  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson F., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lighning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,  
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Du Z.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U02830; AAC48280.1; -.  
DR InterPro; IPR000718; Peptidase\_M13.  
DR Pfam; PF01431; Peptidase\_M13; 1.  
SQ SEQUENCE 433 AA; 50075 MW; 819E8A1EB676B054 CRC64;

Query Match 56.2% ; Score 45; DB 5; Length 433;



Db 813 RAAQGYVETIFGRRL 828  
 |||:|||| |||:

RESULT 5  
 Q9QAR7 PRELIMINARY; PRT; 207 AA.  
 AC Q9QAR7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE I PROTEIN.  
 GN I.

OS Bovine coronavirus (strain LY-138).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11131;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LY-138;  
 RX MEDLINE=98451808; PubMed=9778786;  
 RA Choulien V.N., Kousoulas K.G., Lin X., Storz J.;  
 RT "Nucleotide and predicted amino acid sequences of all genes encoded by  
 the 3' genomic portion (9.5 kb) of respiratory bovine coronaviruses  
 and comparisons among respiratory and enteric coronaviruses.";  
 RL Virus Genes 17:33-42(1998).  
 DR EMBL: AF058943; AAF25516.1; -.  
 SQ SEQUENCE 207 AA; 23094 MW; D354C47342CC6EF1 CRC64;

Query Match 50.0%; Score 40; DB 12; Length 207;  
 Best Local Similarity 50.0%; Pred. No. 19;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRV 16  
 | |||:| | | | |  
 Db 150 RLFLGRLRLYSLRV 165

RESULT 6  
 Q9QAR0 PRELIMINARY; PRT; 207 AA.  
 AC Q9QAR0;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DE I PROTEIN.  
 GN I.

OS Bovine coronavirus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11128;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LSU-94LSS-051-2;  
 RX MEDLINE=98451808; PubMed=9778786;  
 RA Choulien V.N., Kousoulas K.G., Lin X., Storz J.;  
 RT "Nucleotide and predicted amino acid sequences of all genes encoded by  
 the 3' genomic portion (9.5 kb) of respiratory bovine coronaviruses  
 and comparisons among respiratory and enteric coronaviruses.";  
 RL Virus Genes 17:33-42(1998).  
 DR EMBL: AF058943; AAF25516.1; -.  
 SQ SEQUENCE 207 AA; 23012 MW; C84EIF7612C9C3DC CRC64;

Query Match 50.0%; Score 40; DB 12; Length 207;  
 Best Local Similarity 50.0%; Pred. No. 19;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRV 16  
 | |||:| | | | |  
 Db 150 RLFLGRLRLYSLRV 165

RESULT 7

Q9QAQ4 PRELIMINARY; PRT; 207 AA.  
 AC Q9QAQ4;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE I PROTEIN.

GN I.  
 OS Bovine coronavirus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11128;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OK-0514-3;  
 RX MEDLINE=98451808; PubMed=9778786;  
 RA Choulien V.N., Kousoulas K.G., Lin X., Storz J.;  
 RT "Nucleotide and predicted amino acid sequences of all genes encoded by  
 the 3' genomic portion (9.5 kb) of respiratory bovine coronaviruses  
 and comparisons among respiratory and enteric coronaviruses.";  
 RL Virus Genes 17:33-42(1998).  
 DR EMBL: AF058944; AAF25526.1; -.  
 SQ SEQUENCE 207 AA; 22989 MW; C83903664387E059 CRC64;

Query Match 50.0%; Score 40; DB 12; Length 207;  
 Best Local Similarity 50.0%; Pred. No. 19;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RAARLGYRXXRFGXRV 16  
 | |||:| | | | |  
 Db 150 RLFLGRLRLYSLRV 165

RESULT 8

Q9K4V0 PRELIMINARY; PRT; 513 AA.  
 AC Q9K4V0;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE NITRIC OXIDE REDUCTASE REGULATOR NORR.  
 GN NORR.  
 OS Alkaligenes eutrophus (Ralstonia eutropha).  
 OG Plasmid pHG1.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=510;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H16;  
 RA Pohlmann A., Cramm R., Schmelz K., Friedrich B.;  
 RT "A novel NO-respondering regulator controls the reduction of nitric  
 oxide in Ralstonia eutropha.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING  
 DOMAIN.

DR EMBL: AJ278371; CAC00710.1; -.  
 DR InterPro; IPR003593; AAA.  
 DR InterPro; IPR003018; GAF.  
 DR InterPro; IPR002197; HTH\_Fis.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR Pfam; PF01590; GAF; 1.  
 DR Pfam; PF00158; sigma54; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR SMART; SM00065; GAF; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
 DR PROSITE; PS00676; SIGMA54\_INTERACT\_2; 1.  
 DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; 1.  
 DR PROSITE; PS50045; SIGMA54\_INTERACT\_4; 1.  
 KW ATP-binding; DNA-binding; Plasmid; Transcription regulation.



```

Query Match      48.8%; Score 39; DB 5; Length 127;
Best Local Similarity 52.9%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVG 17
   | | | | | | | |
Db 83 RFRMFGKRFMRFGRSVG 99

RESULT 12
QI6964 PRELIMINARY; PRT; 373 AA.
AC QI6964;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FMRFAMIDE (FRAGMENT).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea;
OC Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87132918; PubMed=3816495;
RA Taussig R., Scheller R.H.;
RT "The Aplysia FMRFamide gene encodes sequences related to mammalian
RT brain peptides.";
RL DNA 5:453-461(1986).
DR EMBL; M14960; AAA27752.1; -.
DR InterPro: IPR002544; FARP.
DR Pfam; PF01581; FARP; 23.
FT NON_TER 1
SQ SEQUENCE 373 AA; 43453 MW; 9A309C1514EA8276 CRC64;

Query Match      48.8%; Score 39; DB 5; Length 373;
Best Local Similarity 52.9%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVG 17
   | | | | | | | |
Db 329 RFRMFGKRFMRFGRSVG 345

RESULT 13
O68994 PRELIMINARY; PRT; 415 AA.
AC O68994;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE TRANSPOSASE.
OS Mycobacterium avium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=2151;
RA Eckstein T.M., Silbaq F.S., Inamine I.M., Belisle J.T.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2151;
RA Eckstein T.M., Lambert M.L., Brennan P.J., Belisle J.T., Inamine J.M.;
RT "Identification of a gene cluster involved in glycopeptidolipid
RT biosynthesis and of a gene cluster encoding daunorubicin resistance in
RT two strains of Mycobacterium avium serovar 2.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF060182; AAC71696.1; -.
DR EMBL; AF143772; AAD44226.1; -.

```

```

DR EMBL; AF143772; AAD44203.1; -.
DR InterPro: IPR01207; Transpo_mutator.
DR Pfam; PF00872; Transpo_mutator; 1.
SQ SEQUENCE 415 AA; 45256 MW; 50FF161AA40CA891 CRC64;

Query Match      47.5%; Score 38; DB 2; Length 415;
Best Local Similarity 52.9%; Pred. No. 90;
Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRVG 17
   | | | | | | | |
Db 57 RAQRNGYRHRGLDTRVG 73

RESULT 14
O9CCS8 PRELIMINARY; PRT; 421 AA.
AC O9CCS8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN ML0514.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL583918; CAC30022.1; -.
KW Complete proteome.
SQ SEQUENCE 421 AA; 45625 MW; B25B500D075716B6 CRC64;

Query Match      47.5%; Score 38; DB 2; Length 421;
Best Local Similarity 43.8%; Pred. No. 92;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 RAARLGYRXXRFGXRV 16
   | | | | | | | |
Db 32 RLAKSGYRCRRFADKI 47

RESULT 15
O68139 PRELIMINARY; PRT; 194 AA.
AC O68139;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 21.3 KDA PROTEIN.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003;
RX MEDLINE=9740404; PubMed=9256491;

```

RA Vlcek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;  
RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter  
capsulatus SB1003.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
DR EMBL; AF010496; AAC16229.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 194 AA; 21263 MW; 73F67C4F9F8A9677 CRC64;

Query Match 46.9%; Score 37.5; DB 2; Length 194;  
Best Local Similarity 47.4%; Pred. No. 48;  
Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 1 RAAR---LGYRXXRFGXRV I6  
I::I I: III I:  
Db 147 RSARYGLGWLRLAREGARI 165

Search completed: February 12, 2002, 12:38:39  
Job time: 752 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:31 ; Search time 242.57 Seconds  
(without alignments)  
3.054 Million cell updates/sec

Title: US-09-485-571-22  
Perfect score: 49  
Sequence: 1 YRRFVSYSVR 10

Scoring Table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Length	ID	Description
1	49	100.0	10	20	AAW99409
2	42	85.7	15	20	AAW99411
3	37	75.5	18	16	AAW78776
4	37	75.5	18	20	AAW99403
5	37	75.5	18	21	AAW93616
6	34	69.4	196	22	AAW40747
7	34	69.4	212	21	AAG58075
8	34	69.4	219	21	AAG58074
9	34	69.4	474	21	AAW68773
10	33	67.3	94	21	AAW94288
11	32	65.3	13	19	AAW52500
					Protegrin derivati
					Protegrin derivati
					Protegrin peptide
					Protegrin peptide
					Peptide which may
					Human polypeptide
					Arabidopsis thalia
					Arabidopsis thalia
					Amino acid sequenc
					Soybean glutamyl-t
					Cyclic peptide of

12	65.3	13	19	AAW52513	Cyclic peptide of
13	65.3	13	19	AAW52517	Cyclic peptide of
14	65.3	13	20	AAW17324	Antimicrobial cycl
15	65.3	13	20	AAW17307	Antimicrobial cycl
16	65.3	13	20	AAW17320	Antimicrobial cycl
17	65.3	14	18	AAW36220	Antimicrobial prot
18	65.3	14	19	AAW52516	Cyclic peptide of
19	65.3	14	19	AAW52567	Cyclic peptide of
20	65.3	14	19	AAW52568	Cyclic peptide of
21	65.3	14	19	AAW52540	Cyclic peptide of
22	65.3	14	19	AAW52536	Cyclic peptide of
23	65.3	14	19	AAW52537	Cyclic peptide of
24	65.3	14	20	AAW17347	Antimicrobial cycl
25	65.3	14	20	AAW17343	Antimicrobial cycl
26	65.3	14	20	AAW17344	Antimicrobial cycl
27	65.3	14	20	AAW17323	Antimicrobial cycl
28	65.3	14	20	AAW17375	Antimicrobial cycl
29	65.3	17	18	AAW09078	Cationic, antimicr
30	65.3	190	20	AAW36971	Amino acid sequenc
31	65.3	412	22	AAW81373	Human AFP protein
32	65.3	772	22	AAW92872	Human protein sequ
33	63.3	14	19	AAW52504	Cyclic peptide of
34	63.3	14	20	AAW17311	Antimicrobial cycl
35	63.3	18	18	AAW18151	Cationic, antimicr
36	63.3	113	21	AAW10631	Human VEGF-X prote
37	63.3	113	21	AAW10632	Human VEGF-X prote
38	63.3	121	22	AAW74034	Synthetic protein
39	63.3	123	22	AAW74033	Synthetic protein
40	63.3	124	22	AAW52459	Mycobacterium tube
41	63.3	149	21	AAW10642	Human VEGF-X PDGF-
42	63.3	227	21	AAW10637	Human VEGF-X prote
43	63.3	227	21	AAW10638	Human VEGF-X prote
44	63.3	251	20	AAW36790	Amino acid sequenc
45	63.3	265	21	AAW41354	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1  
AAW99409  
ID AAW99409 standard; peptide; 10 AA.  
XX  
AC AAW99409:  
DT 08-JUN-1999 (first entry)  
XX  
DE Protegrin derivative peptide SM2194.  
XX  
Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

XX Synthetic.  
OS  
PN WO9907728-A2.  
XX  
PD 18-FEB-1999.

XX 06-AUG-1998; 98WO-FR01757.  
XX  
XX 12-AUG-1997; 97FR-0010297.  
XX  
XX (SYNT-) SYNT:EM SA.

XX Calas B, Chavanieu A, Grassy G, Kaczorek M;  
XX WPI; 1959-190034/16.

XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
PT as carriers to deliver active agents into cells  
XX

PS Claim 7; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 10 AA;

Query Match 100.0%; Score 49; DB 20; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0017;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YRRFVSVR 10

Db 1 YRRFVSVR 10

RESULT 2

AAW99411  
 ID AAW99411 standard; peptide; 15 AA.

XX

AC AAW99411;

XX 08-JUN-1999 (first entry)

XX Protegrin derivative peptide SM2193.

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX Synthetic.

XX WO9907728-A2.

XX 18-FEB-1999.

XX 06-AUG-1998; 98WO-FR01757.

XX 12-AUG-1997; 97FR-0010297.

XX (SYNT-) SYNT:EM SA.

XX Calas B, Chavanleu A, Grassy G, Kaczorek M;

XX WPI; 1999-190034/16.

XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX Claim 7; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 15 AA;

Query Match 85.7%; Score 42; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.063;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 RRRFVSVR 10

Db 7 RRRFVSVR 15

RESULT 3

AAW78776  
 ID AAW78776 standard; peptide; 18 AA.

XX

AC AAW78776;

XX 08-OCT-1995 (first entry)

XX Protegrin peptide sequence.

XX protegrin; antibiotic; antimicrobial; antiviral; antibacterial;  
 KW antifungal;

XX Synthetic.

XX WO9503325-A.

XX 02-FEB-1995.

XX 20-JUL-1994; 94WO-US08305.

XX 20-JUL-1993; 93US-0093926.

XX 26-JUL-1993; 93US-0095769.

XX 13-JAN-1994; 94US-0182483.

XX 17-MAY-1994; 94US-0243879.

XX (REGC ) UNIV CALIFORNIA.

XX Harwig SSL, Kokryakov VN, Lehrer RL;

XX WPI; 1995-075188/10.

XX Antibiotic peptide-based cpds. designated protegrin(s) - are  
 PT useful for treating and preventing viral and microbial infections  
 PT and as preservatives

XX Disclosure; Page 19; 80pp; English.

XX New peptides are disclosed which are designated "protegrins". The  
 CC peptides are useful as antibacterial, antiviral and antifungal agents in  
 CC both animals and plants. The peptides are 16-18 amino acids in length  
 CC and are characterised by four invariant Cys residues at positions 6, 8,  
 CC 13 and 15 and either (1) by a characteristic pattern of basic and  
 CC hydrophobic amino acids and/or (2) being isolatable from animal (e.g.  
 CC porcine) leukocytes; or analogues of these peptides in which 1-4 of the  
 CC Cys residues is/are replaced by hydrophobic or small amino acids. The  
 CC peptides can be produced synthetically and some can be produced  
 CC recombinantly or can be isolated and purified from their native sources.  
 CC The peptides can be modified by N-acylation and/or C-terminal amidation  
 CC or esterification, and can be in linear or cystine-bridged form. D-Amino  
 CC acid residues can be present.  
 CC The present sequence is a specific example of the protegrin  
 CC analogues in which Cys(6, 8, 13, 15) have all been replaced by Ser.

XX Sequence 18 AA;

Query Match 75.5%; Score 37; DB 16; Length 18;

Best Local Similarity 100.0%; Pred. No. 0.72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 2 RRRFSVSV 9  
 Db 9 rrrfsvsv 16

RESULT 4  
 AAW9403  
 ID AAW9403 standard; peptide; 18 AA.  
 XX  
 AC AAW9403;  
 XX  
 DT 08-JUN-1999 (first entry)  
 XX  
 DE Protegrin derivative peptide SM1738.  
 XX  
 KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.  
 XX  
 OS Synthetic.  
 XX  
 XX WO9907728-A2.  
 PN  
 PD 18-FEB-1999.  
 XX  
 PF 06-AUG-1998; 98WO-FR01757.  
 XX  
 PR 12-AUG-1997; 97FR-0010297.  
 XX  
 PA (SYNT-) SYNT:EM SA.  
 XX  
 XX Calas B, Chavanieu A, Grassy G, Kaczorek M;  
 PI  
 XX WPI; 1999-190034/16.  
 DR  
 XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells  
 XX  
 XX Claim 7; Page 28; 37pp; French.  
 PS  
 CC This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.  
 XX  
 SQ Sequence 18 AA;

Query Match 75.5%; Score 37; DB 20; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 0.72; Mismatches 0; Gaps 0;  
 Matches 8; Conservative 0; Indels 0; Indels 0; Gaps 0;

QY 2 RRRFSVSV 9  
 Db 9 rrrfsvsv 16

RESULT 5  
 AAY93616  
 ID AAY93616 standard; peptide; 18 AA.  
 XX  
 AC AAY93616;  
 XX  
 DT 25-SEP-2000 (first entry)  
 XX

XX  
 DE Peptide which may be linked to anticancer agents.  
 XX  
 KW Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 KW cancer.  
 XX  
 OS Unident.:filed.  
 XX  
 PN WO200032237-A1.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 26-NOV-1999; 99WO-FR02939.  
 XX  
 PR 30-NOV-1998; 98FR-0015073.  
 XX  
 PA (SYNT-) SYNT:EM SA.  
 XX  
 PI Tamsamani J, Kaczorek M, Colin De Verdiere A;  
 XX  
 DR WPI; 2000-412166/35.  
 XX  
 XX New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells  
 XX  
 PS Disclosure; Page 8; 34pp; French.  
 XX  
 CC The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.  
 XX  
 SQ Sequence 18 AA;

Query Match 75.5%; Score 37; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 0.72; Mismatches 0; Gaps 0;  
 Matches 8; Conservative 0; Indels 0; Indels 0; Gaps 0;

QY 2 RRRFSVSV 9  
 Db 9 rrrfsvsv 16

RESULT 6  
 AAM40747  
 ID AAM40747 standard; Protein; 196 AA.  
 XX  
 AC AAM40747;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 5678.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX

PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
DR N-PSDB; AAI59903.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 2; SEQ ID NO 5678; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAI38642-AAI42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, cancer diagnosis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 196 AA;

Query Match 69.4%; Score 34; DB 22; Length 196;  
Best Local Similarity 66.7%; Pred. No. 35;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRFVSIV 9  
Db ::|||:|  
72 fkrfslsv 80

RESULT 7  
AAG58075  
ID AAG58075 standard; Protein; 212 AA.  
XX  
AC AAG58075;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 74922.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX

XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135153.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.

```
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161320.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 59.4%; Score 34; DB 21; Length 212;
Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RRRFSVSV 9
Db 60 rrrfsvsm 67

RESULT 8
AAG58074
ID AAG58074 standard; Protein; 219 AA.
XX AAG58074;
AC
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 74921.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
```

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 69.4%; Score 34; DB 21; Length 219;  
 Best Local Similarity 87.5%; Pred. No. 40;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RRRFSVS 9  
 |||||  
 Db 67 rrrfsvsm 74

RESULT 9  
 AAY68773  
 ID AAY68773 standard; Protein: 474 AA.

AC AAY68773;

XX 16-MAY-2000 (first entry)

XX Amino acid sequence of a human phosphorylation effector PHSP-5.  
 DE Human; phosphorylation effector; PHSP; proliferative disorder;  
 KW Immune disorder; neuronal disorder.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 14  
 FT Modified-site 22 /note= "potential phosphorylation site"  
 FT Modified-site 26 /note= "potential phosphorylation site"  
 FT Modified-site 26 /note= "potential phosphorylation site"  
 FT Modified-site 62 /note= "potential phosphorylation site"  
 FT Modified-site 66 /note= "potential phosphorylation site"  
 FT Modified-site 89 /note= "potential phosphorylation site"  
 FT Modified-site 94 /note= "potential phosphorylation site"  
 FT Modified-site 98 /note= "potential phosphorylation site"

FT Modified-site 128 /note= "potential phosphorylation site"  
 FT Modified-site 132 /note= "potential phosphorylation site"  
 FT Modified-site 155 /note= "potential phosphorylation site"  
 FT Region 144..425 /note= "potential phosphorylation site"  
 FT Modified-site 204 /note= "protein kinase family signature sequence"  
 FT Modified-site 211 /note= "potential phosphorylation site"  
 FT Modified-site 320 /note= "potential phosphorylation site"  
 FT Modified-site 336 /note= "potential phosphorylation site"  
 FT Modified-site 345 /note= "potential phosphorylation site"  
 FT Modified-site 359 /note= "potential phosphorylation site"  
 FT Modified-site 427 /note= "potential phosphorylation site"  
 FT Modified-site 443 /note= "potential phosphorylation site"  
 FT Modified-site 443 /note= "potential phosphorylation site"

WO200006728-A2.

PD 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17132.

PR 28-JUL-1998; 98US-0123494.

PR 14-OCT-1998; 98US-0152814.

PR 03-NOV-1998; 98US-0173482.

PR 19-NOV-1998; 98US-0106889.

PR 22-DEC-1998; 98US-0109093.

PR 12-JAN-1999; 98US-0113796.

PR 12-JAN-1999; 98US-0173482.

XX 12-JAN-1999; 99US-0229005.

XX (INCY-) INCYTE PHARM INC.

PA Hillman JL, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;

PI Patterson C, Bandman O, Au-Young J, Gorgone GA, Yue H, Azimzai Y;

PI Reddy R, Lu DAM, Shih LL;

XX WPI: 2000-183135/16.

DR N-PSDB; AA246142.

XX New human phosphorylation effectors useful for the diagnosis, treatment

PT and prevention of proliferative, immune and neuronal disorders -

XX Claim 1: Page 83-84; 142pp; English.

XX AAY68769-95 and AAY68797-99 represent human phosphorylation effectors

CC (PHSP), designated PHSP1-PHSP31 (the protein sequence for PHSP28 is not

CC given in the specification). The sequences were isolated from CDNA

CC libraries prepared from various human tissues. The PHSP proteins are

CC useful for the diagnosis, treatment and prevention of proliferative

CC disorders, immune disorders and neuronal disorders. The PHSP proteins

CC form pharmaceutical compositions which useful for treating or preventing

CC disorders associated with decreased PHSP expression/activity. PHSP

CC antagonists are useful for treating or preventing disorders associated

CC with increased PHSP expression/activity.

XX Sequence: 474 AA;

Query Match 69.4%; Score 34; DB 21; Length 474;  
 Best Local Similarity 87.5%; Pred. No. 40;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRFSVS 9  
 Db 9 fkrfsls 17

RESULT 10  
 AAY94288  
 ID AAY94288 standard; Protein; 94 AA.  
 XX  
 AC AAY94288;  
 XX  
 DT 10-AUG-2000 (first entry)  
 XX  
 DE Soybean glutamyl-tRNA synthetase # 2.  
 KW Soybean; glutamyl-tRNA synthetase; aminoacyl-tRNA synthetase;  
 KW AARS; herbicide; plant toxin; protein synthesis inhibition; enzyme.  
 XX  
 OS Glycine max.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 63  
 FT /label= unknown  
 FT /note= "Encoded by TCCA"  
 FT Misc-difference 83  
 FT /note= "Encoded by GATC"  
 XX  
 XX WO200028057-A2.  
 PN  
 XX  
 XX 18-MAY-2000.  
 PD  
 XX  
 XX 09-NOV-1999; 99WO-US26478.  
 PF  
 XX  
 XX 10-NOV-1998; 98US-0107789.  
 PR  
 XX  
 XX (DUPO ) DU PONT DE NEMOURS & CO E. I.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 XX Famodu OO, Simmons C;  
 PI  
 XX  
 XX WPI; 2000-387421/33.  
 DR  
 XX  
 XX N-PSDB; AAA27419.  
 DR  
 XX  
 XX New polynucleotide encoding plant aminoacyl-tRNA synthase and the  
 PT encoded polypeptide, useful e.g. for regulating gene expression -  
 PT  
 XX  
 PS Claim 42; Page 71; 74pp; English.  
 CC  
 CC The present sequence is the soybean glutamyl-tRNA synthetase. This  
 CC enzyme is an aminoacyl-tRNA synthetase (AARS). AARSs charge (acylate)  
 CC specific tRNAs with amino acids for use in protein synthesis. Glutamyl-  
 CC tRNA synthetase therefore charges an glutamyl-tRNA with glutamate. Since  
 CC this enzyme has a crucial role in protein synthesis and therefore life,  
 CC any agent that inhibits or disrupts protein synthesis is likely to be  
 CC toxic. The present sequence could therefore be used as a basis for  
 CC testing whether the encoded aminoacyl-tRNA synthetase is sensitive to  
 CC known inhibitors or other chemicals and hence could be used in the  
 CC discovery of potential herbicides.  
 XX  
 XX Sequence 94 AA;

Query Match 67.3%; Score 33; DB 21; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RRRFSVS 8  
 Db 2 rrrfsvs 8

RESULT 11  
 AAW52500

ID AAW52500 standard; peptide; 13 AA.  
 XX  
 AC AAW52500;  
 XX  
 DT 01-JUL-1998 (first entry)  
 XX  
 DE Cyclic peptide of the invention.  
 XX  
 KW Loop region; cyclic peptide; antimicrobial; disinfectant; therapy;  
 KW preservative; amphipathic anti-parallel beta-sheet region; plant disease.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9803192-A1.  
 XX  
 XX 29-JAN-1998.  
 PD  
 XX  
 XX 23-JUL-1997; 97WO-US12974.  
 PF  
 XX  
 XX 24-JUL-1996; 96US-0685589.  
 PR  
 XX  
 XX (INTR-) INTRABIOTICS PHARM INC.  
 PA  
 XX  
 XX Chang C, Chen J, . Gu L;  
 PI  
 XX  
 XX WPI; 1998-120472/11.  
 DR  
 XX  
 XX New cyclic peptide(s) with antimicrobial activity - contain  
 PT amphipathic beta-sheet, loop and beta-turn regions, have better  
 PT activity, bio:availability and protease resistance than linear  
 PT analogues  
 XX  
 XX Claim 11; Page 155; 160pp; English.  
 XX  
 CC This sequence is an example of a cyclic peptides (I) of the invention,  
 CC which has: (a) an amphipathic anti-parallel beta-sheet region (SR), a  
 CC loop region (LR) and a beta-turn region (TR); (b) a net positive charge  
 CC at physiological pH; and (c) at least one basic amino acid (aa) in LR or  
 CC TR. (I) are broad spectrum antimicrobials, specifically for use against  
 CC E. coli, Pseudomonas aeruginosa, methicillin-resistant Staphylococcus  
 CC aureus (MRSA), vancomycin-resistant Enterococcus faecium and  
 CC penicillin-resistant Streptococcus pneumoniae. More generally they are  
 CC active against gram-positive or -negative bacteria, fungi, yeast and  
 CC protozoa. Apart from clinical uses, (I) are also used as disinfectants  
 CC and preservatives for medical equipment, foods, cosmetics etc., also for  
 CC treatment of plant diseases. Compared with non-cyclised analogues (i.e.  
 CC tachyplesin and protegrin type peptides), (I) and are more effective,  
 CC with better bioavailability and/or serum half-life (increased resistance  
 CC to proteolysis). They are more suitable for oral administration, can be  
 CC used at lower doses and are unlikely to induce development of resistant  
 CC strains.  
 XX  
 XX Sequence 13 AA;

Query Match 65.3%; Score 32; DB 19; Length 13;  
 Best Local Similarity 77.8%; Pred. No. 4.9;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVS 10  
 Db 4 rrrfcvvr 12

RESULT 12  
 AAW52513  
 ID AAW52513 standard; peptide; 13 AA.  
 XX  
 AC AAW52513;  
 XX  
 DT 01-JUL-1998 (first entry)  
 XX  
 DE Cyclic peptide of the invention.

XX KW Loop region: cyclic peptide; antimicrobial; disinfectant; therapy;  
 XX KW preservative; amphipathic anti-parallel beta-sheet region; plant disease.  
 XX OS Synthetic.  
 XX FH Key Location/Qualifiers  
 FT Modified-site 13  
 FT /note= "Cyclohexylalanine"  
 XX XX  
 XX PN WO9803192-A1.  
 XX PD 29-JAN-1998.  
 XX PF 23-JUL-1997; 97WO-US12974.  
 XX PR 24-JUL-1996; 96US-0685589.  
 XX PA (INTR-) INFRABIOTICS PHARM INC.  
 XX PI Chang C, Chen J, Gu L;  
 XX WI; 1998-120472/11.  
 XX DR  
 XX PT New cyclic peptide(s) with antimicrobial activity - contain  
 PT amphipathic beta-sheet, loop and beta-turn regions, have better  
 PT activity, bio:availability and protease resistance than linear  
 PT analogues  
 XX XX  
 XX PS Claim 11; Page 155; 160pp; English.  
 XX CC This sequence is an example of a cyclic peptides (I) of the invention,  
 CC which has: (a) an amphipathic anti-parallel beta-sheet region (SR), a  
 CC loop region (LR) and a beta-turn region (TR); (b) a net positive charge  
 CC at physiological pH; and (c) at least one basic amino acid (aa) in LR or  
 CC TR. (I) are broad spectrum antimicrobials, specifically for use against  
 CC E. coli, Pseudomonas aeruginosa, methicillin-resistant Staphylococcus  
 CC aureus (MRSA), vancomycin-resistant Enterococcus faecium and  
 CC penicillin-resistant Streptococcus pneumoniae. More generally they are  
 CC active against Gram-positive or -negative bacteria, fungi, yeast and  
 CC protozoa. Apart from clinical uses, (I) are also used as disinfectants  
 CC and preservatives for medical equipment, foods, cosmetics etc., also for  
 CC treatment of plant diseases. Compared with non-cyclised analogues (i.e.  
 CC tachyplesin and protegrin type peptides), (I) and are more effective,  
 CC with better bioavailability and/or serum half-life (increased resistance  
 CC to proteolysis). They are more suitable for oral administration, can be  
 CC used at lower doses and are unlikely to induce development of resistant  
 CC strains.  
 XX SQ Sequence 13 AA;  
 Query Match 65.3%; Score 32; DB 19; Length 13;  
 Best Local Similarity 77.8%; Pred. No. 4.9;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 RRRFSVSVR 10  
 Db 4 rrrfcvcr 12  
 RESULT 13  
 AAW52517  
 ID AAW52517 standard; peptide; 13 AA.  
 XX AC AAW52517;  
 XX DT 01-JUL-1998 (first entry)  
 XX DE Cyclic peptide of the invention.  
 XX KW Loop region: cyclic peptide; antimicrobial; disinfectant; therapy;  
 KW KW preservative; amphipathic anti-parallel beta-sheet region; plant disease.

XX OS Synthetic.  
 XX PN WO9803192-A1.  
 XX PD 29-JAN-1998.  
 XX PF 23-JUL-1997; 97WO-US12974.  
 XX PR 24-JUL-1996; 96US-0685589.  
 XX PA (INTR-) INFRABIOTICS PHARM INC.  
 XX PI Chang C, Chen J, Gu L;  
 XX WI; 1998-120472/11.  
 XX DR  
 XX PT New cyclic peptide(s) with antimicrobial activity - contain  
 PT amphipathic beta-sheet, loop and beta-turn regions, have better  
 PT activity, bio:availability and protease resistance than linear  
 PT analogues  
 XX XX  
 XX PS Claim 11; Page 155; 160pp; English.  
 XX CC This sequence is an example of a cyclic peptides (I) of the invention,  
 CC which has: (a) an amphipathic anti-parallel beta-sheet region (SR), a  
 CC loop region (LR) and a beta-turn region (TR); (b) a net positive charge  
 CC at physiological pH; and (c) at least one basic amino acid (aa) in LR or  
 CC TR. (I) are broad spectrum antimicrobials, specifically for use against  
 CC E. coli, Pseudomonas aeruginosa, methicillin-resistant Staphylococcus  
 CC aureus (MRSA), vancomycin-resistant Enterococcus faecium and  
 CC penicillin-resistant Streptococcus pneumoniae. More generally they are  
 CC active against Gram-positive or -negative bacteria, fungi, yeast and  
 CC protozoa. Apart from clinical uses, (I) are also used as disinfectants  
 CC and preservatives for medical equipment, foods, cosmetics etc., also for  
 CC treatment of plant diseases. Compared with non-cyclised analogues (i.e.  
 CC tachyplesin and protegrin type peptides), (I) and are more effective,  
 CC with better bioavailability and/or serum half-life (increased resistance  
 CC to proteolysis). They are more suitable for oral administration, can be  
 CC used at lower doses and are unlikely to induce development of resistant  
 CC strains.  
 XX SQ Sequence 13 AA;  
 Query Match 65.3%; Score 32; DB 19; Length 13;  
 Best Local Similarity 77.8%; Pred. No. 4.9;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 RRRFSVSVR 10  
 Db 4 rrrfcvcr 12  
 RESULT 14  
 AAY17324  
 ID AAY17324 standard; peptide; 13 AA.  
 XX AC AAY17324;  
 XX DT 21-JUL-1999 (first entry)  
 XX DE Antimicrobial cyclic peptide SEQ ID NO:171.  
 XX KW Cyclic peptide; antimicrobial; amphiphilic beta-sheet; biostatic;  
 KW biocidal; bacteria; yeast; Candida species; fungi; protozoa;  
 KW Escherichia coli; Pseudomonas aeruginosa; infection; preservative;  
 KW vancomycin-resistant Enterococcus faecium; disinfectant; food;  
 KW methicillin-resistant Staphylococcus aureus; medical equipment;  
 KW penicillin-resistant Streptococcus pneumoniae; cosmetic.  
 XX OS Synthetic.

PN W09921879-A1.  
 XX  
 PD 06-MAY-1999.  
 XX  
 PF 27-OCT-1997; 97WO-US19557.  
 XX  
 PR 27-OCT-1997; 97WO-US19557.  
 XX  
 PA (INTR-) INTRABIOTICS PHARM INC.  
 XX  
 PI Chang C, Chen J, Gu L;  
 XX  
 DR WPI; 1999-312941/26.  
 XX  
 PT Antimicrobial cyclic peptides with amphiphilic beta-sheet region  
 XX  
 PS Claim 11; Page 157; 167pp; English.  
 XX  
 CC The present invention describes cyclic peptides (I): (i) comprising an  
 CC amphiphilic anti-parallel beta-sheet region (A), a loop region (B) and  
 CC a beta-turn (C); (ii) having net positive charge at physiological pH;  
 CC and (iii) including at least one basic amino acid (aa) in (B) or (C).  
 CC AAY09554 to AAY09683 and AAY17301 to AAY17385 represent specifically  
 CC claimed cyclic peptides and peptide segments. (I) are broad-spectrum  
 CC antimicrobials (biostatic or biocidal), effective against bacteria,  
 CC yeast (e.g. Candida species), fungi and protozoa. Particularly they  
 CC are used to control Enterococcus faecium, methicillin-resistant  
 CC Staphylococcus aureus and penicillin-resistant Streptococcus  
 CC pneumoniae. (A) are used to treat or prevent infections, in animals or  
 CC plants, also as preservatives and disinfectants for medical equipment,  
 CC foods, cosmetics, optionally as mixtures or in combination with other  
 CC antimicrobials. Compared with linear analogues, (I) are more effective  
 CC with better bioavailability and/or serum half-life (better resistance  
 CC to proteolysis, allowing lower doses and making them more suitable for  
 CC oral delivery). Since (I) are structurally related to naturally  
 CC occurring antimicrobial peptides, they are less likely to induce  
 CC development of resistant strains.  
 XX  
 SQ Sequence 13 AA;  
 Query Match 65.3%; Score 32; DB 20; Length 13;  
 Best Local Similarity 77.8%; Pred. No. 4.9;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 RRRFSVSVR 10  
 DB 4 rrrfcvvr 12  
 RESULT 15  
 ID AAY17307 standard; Peptide; 13 AA.  
 AC AAY17307;  
 DT 21-JUL-1999 (first entry)  
 DE Antimicrobial cyclic peptide SEQ ID NO:154.  
 XX  
 KW Cyclic peptide; antimicrobial; amphiphilic beta-sheet; biostatic;  
 KW biocidal; bacteria; yeast; Candida species; fungi; protozoa;  
 KW Escherichia coli; Pseudomonas aeruginosa; infection; preservative;  
 KW vancomycin-resistant Enterococcus faecium; disinfectant; food;  
 KW methicillin-resistant Staphylococcus aureus; medical equipment;  
 KW penicillin-resistant Streptococcus pneumoniae; cosmetic.  
 XX  
 OS Synthetic.  
 XX  
 PN W09921879-A1.  
 XX  
 PD 06-MAY-1999.

XX  
 PF 27-OCT-1997; 97WO-US19557.  
 XX  
 PR 27-OCT-1997; 97WO-US19557.  
 XX  
 PA (INTR-) INTRABIOTICS PHARM INC.  
 XX  
 PI Chang C, Chen J, Gu L;  
 XX  
 DR WPI; 1999-312941/26.  
 XX  
 PT Antimicrobial cyclic peptides with amphiphilic beta-sheet region  
 XX  
 PS Claim 11; Page 156; 167pp; English.  
 XX  
 CC The present invention describes cyclic peptides (I): (i) comprising an  
 CC amphiphilic anti-parallel beta-sheet region (A), a loop region (B) and  
 CC a beta-turn (C); (ii) having net positive charge at physiological pH;  
 CC and (iii) including at least one basic amino acid (aa) in (B) or (C).  
 CC AAY09554 to AAY09683 and AAY17301 to AAY17385 represent specifically  
 CC claimed cyclic peptides and peptide segments. (I) are broad-spectrum  
 CC antimicrobials (biostatic or biocidal), effective against bacteria,  
 CC yeast (e.g. Candida species), fungi and protozoa. Particularly they  
 CC are used to control Enterococcus faecium, methicillin-resistant  
 CC Staphylococcus aureus and penicillin-resistant Streptococcus  
 CC pneumoniae. (A) are used to treat or prevent infections, in animals or  
 CC plants, also as preservatives and disinfectants for medical equipment,  
 CC foods, cosmetics, optionally as mixtures or in combination with other  
 CC antimicrobials. Compared with linear analogues, (I) are more effective  
 CC with better bioavailability and/or serum half-life (better resistance  
 CC to proteolysis, allowing lower doses and making them more suitable for  
 CC oral delivery). Since (I) are structurally related to naturally  
 CC occurring antimicrobial peptides, they are less likely to induce  
 CC development of resistant strains.  
 XX  
 SQ Sequence 13 AA;  
 Query Match 65.3%; Score 32; DB 20; Length 13;  
 Best Local Similarity 77.8%; Pred. No. 4.9;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 RRRFSVSVR 10  
 DB 4 rrrfcvvr 12  
 Search completed: February 12, 2002, 12:30:32  
 Job time: 365 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:23 ; Search time 106.12 Seconds  
(without alignments)  
2.121 Million cell updates/sec

Title: US-09-485-571-22  
Perfect score: 49  
Sequence: 1 YRRFSVSVR 10

Scoring Table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	75.5	18	1 US-08-182-483A-28	Sequence 28, Appl
2	37	75.5	18	1 US-08-243-879A-27	Sequence 27, Appl
3	37	75.5	18	1 US-08-499-523-48	Sequence 48, Appl
4	37	75.5	18	4 US-09-128-345-48	Sequence 48, Appl
5	32	65.3	13	2 US-08-685-589A-154	Sequence 154, Appl
6	32	65.3	13	2 US-08-685-589A-167	Sequence 167, Appl
7	32	65.3	13	2 US-08-685-589A-171	Sequence 171, Appl
8	32	65.3	14	2 US-08-685-589A-170	Sequence 170, Appl
9	32	65.3	14	2 US-08-685-589A-190	Sequence 190, Appl
10	32	65.3	14	2 US-08-685-589A-191	Sequence 191, Appl
11	32	65.3	14	2 US-08-685-589A-194	Sequence 194, Appl
12	32	65.3	14	2 US-08-685-589A-222	Sequence 222, Appl
13	31	63.3	14	2 US-08-685-589A-158	Sequence 158, Appl
14	31	63.3	373	2 US-08-599-171A-26	Sequence 26, Appl
15	31	63.3	373	2 US-08-646-590B-26	Sequence 26, Appl
16	31	63.3	373	3 US-09-069-226-26	Sequence 26, Appl
17	31	63.3	373	4 US-09-412-184-26	Sequence 26, Appl
18	30	61.2	604	1 US-08-487-753-2	Sequence 2, Appli
19	30	61.2	604	2 US-08-480-065-2	Sequence 2, Appli
20	30	61.2	604	3 US-08-487-744-2	Sequence 2, Appli
21	30	61.2	604	5 PCT-US93-09167-2	Sequence 2, Appli
22	30	61.2	808	2 US-08-629-291A-33	Sequence 33, Appl
23	30	61.2	808	2 US-08-658-335B-33	Sequence 33, Appl
24	30	61.2	1257	2 US-08-750-152A-2	Sequence 2, Appli
25	29	59.2	12	2 US-08-752-852A-57	Sequence 57, Appl
26	29	59.2	12	2 US-08-752-852A-59	Sequence 59, Appl
27	29	59.2	12	2 US-08-752-852A-75	Sequence 75, Appl

28	29	59.2	12	2 US-08-752-852A-76	Sequence 76, Appl
29	29	59.2	12	2 US-08-752-852A-92	Sequence 92, Appl
30	29	59.2	12	3 US-08-752-853-24	Sequence 24, Appl
31	29	59.2	13	2 US-08-752-852A-41	Sequence 41, Appl
32	29	59.2	18	1 US-08-499-523-63	Sequence 63, Appl
33	29	59.2	18	1 US-08-499-523-65	Sequence 65, Appl
34	29	59.2	18	1 US-08-499-523-67	Sequence 67, Appl
35	29	59.2	18	2 US-08-752-852A-86	Sequence 86, Appl
36	29	59.2	18	2 US-08-752-852A-230	Sequence 230, Appl
37	29	59.2	18	4 US-09-128-345-63	Sequence 63, Appl
38	29	59.2	18	4 US-09-128-345-65	Sequence 65, Appl
39	29	59.2	18	4 US-09-128-345-67	Sequence 67, Appl
40	29	59.2	18	4 US-09-128-344A-72	Sequence 72, Appl
41	29	59.2	18	4 US-09-128-344A-193	Sequence 193, Appl
42	29	59.2	356	4 US-09-232-191-4	Sequence 4, Appli
43	29	59.2	356	4 US-09-232-200-4	Sequence 4, Appli
44	29	59.2	356	4 US-09-232-197-4	Sequence 4, Appli
45	29	59.2	514	4 US-08-925-230-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-08-182-483A-28  
; Sequence 28, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,483A  
; FILING DATE: 13-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-182-483A-28

Query Match. 75.5%; Score 37; DB 1; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 RRRFSVSV 9  
DB 9 RRRFSVSV 16

US-08-243-879A-27  
; Sequence 27, Application US/08243879A  
; Patent No. 5708145  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: A NEW PROTEGRIN  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington, DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/243,879A  
; FILING DATE: 17-MAY-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-499-523-48

Query Match 75.5%; Score 37; DB 1; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RRRFSVSV 9  
| | | | | | | |  
Db 9 RRRFSVSV 16

RESULT 4  
US-09-128-345-48  
; Sequence 48, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KORRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-0054-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-128-345-48

US-08-243-879A-27  
; Sequence 27, Application US/08243879A  
; Patent No. 5708145  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: A NEW PROTEGRIN  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington, DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/243,879A  
; FILING DATE: 17-MAY-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-243-879A-27

Query Match 75.5%; Score 37; DB 1; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RRRFSVSV 9  
| | | | | | | |  
Db 9 RRRFSVSV 16

RESULT 3  
US-08-499-523-48  
; Sequence 48, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KORRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

Query Match 75.5%; Score 37; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 RRFESVSV 9  
      |||||  
Db 9 RRFESVSV 16

## RESULT 5

US-08-685-589A-154  
; Sequence 154, Application US/08685589A  
; Patent No. 5916872  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Leo  
; APPLICANT: Chen, Jie  
; TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
; SPECTRUM ANTIMICROBIAL ACTIVITY  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,589A  
; FILING DATE: 24-JUL-1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-026-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 154:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: circular  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..13  
; OTHER INFORMATION: /product= "Cyclic"  
US-08-685-589A-154

Query Match 65.3%; Score 32; DB 2; Length 13;  
Best Local Similarity 77.8%; Pred. No. 3.5;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 RRFESVSV 10  
      |||||  
Db 4 RRFECVCR 12

## RESULT 6

US-08-685-589A-167

; Sequence 167, Application US/08685589A  
; Patent No. 5916872  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Leo  
; APPLICANT: Chen, Jie  
; TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
; SPECTRUM ANTIMICROBIAL ACTIVITY  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,589A  
; FILING DATE: 24-JUL-1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-026-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 167:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: circular  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..13  
; OTHER INFORMATION: /product= "Cyclic"  
US-08-685-589A-167

Query Match 65.3%; Score 32; DB 2; Length 13;  
Best Local Similarity 77.8%; Pred. No. 3.5;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 RRFESVSV 10  
      |||||  
Db 4 RRFECVCR 12

## RESULT 7

US-08-685-589A-171  
; Sequence 171, Application US/08685589A  
; Patent No. 5916872  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Leo  
; APPLICANT: Chen, Jie  
; TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
; SPECTRUM ANTIMICROBIAL ACTIVITY

NUMBER OF SEQUENCES: 222  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,589A  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-026-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 171:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..13  
OTHER INFORMATION: /product= "cyclic"  
US-08-685-589A-171

Query Match 65.3%; Score 32; DB 2; Length 13;  
Best Local Similarity 77.8%; Pred. No. 3.5;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSVR 10  
|||||  
Db 4 RRRFCVCR 12

RESULT 8  
US-08-685-589A-170  
Sequence 170, Application US/08685589A  
Patent No. 5916872  
GENERAL INFORMATION:  
APPLICANT: Chang, Conway  
APPLICANT: Gu, Leo  
TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
TITLE OF INVENTION: SPECTRUM ANTIMICROBIAL ACTIVITY  
NUMBER OF SEQUENCES: 222  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,589A  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-026-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 170:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..14  
OTHER INFORMATION: /product= "cyclic"  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /product= "Xaa=Cha=cyclohexylalanine"  
US-08-685-589A-170

Query Match 65.3%; Score 32; DB 2; Length 14;  
Best Local Similarity 77.8%; Pred. No. 3.8;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSVR 10  
|||||  
Db 4 RRRFCVCR 12

RESULT 9  
US-08-685-589A-190  
Sequence 190, Application US/08685589A  
Patent No. 5916872  
GENERAL INFORMATION:  
APPLICANT: Chang, Conway  
APPLICANT: Gu, Leo  
TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
TITLE OF INVENTION: SPECTRUM ANTIMICROBIAL ACTIVITY  
NUMBER OF SEQUENCES: 222  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,589A  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-026-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 190:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..14  
OTHER INFORMATION: /product= "Cyclic"  
US-08-685-589A-190

Query Match 65.3%; Score 32; DB 2; Length 14;  
Best Local Similarity 77.8%; Pred. No. 3.8;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSR 10  
IIII I I I  
Db 4 RRRCVCVR 12

RESULT 10  
US-08-685-589A-191  
Sequence 191, Application US/08685589A  
Patent No. 5916872  
GENERAL INFORMATION:  
APPLICANT: Chang, Conway  
APPLICANT: Gu, Leo  
APPLICANT: Chen, Jie  
TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
TITLE OF INVENTION: SPECTRUM ANTIMICROBIAL ACTIVITY  
NUMBER OF SEQUENCES: 222  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,589A  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-026-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid

STRANDEDNESS: unknown  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..14  
OTHER INFORMATION: /product= "Cyclic"  
US-08-685-589A-191

Query Match 65.3%; Score 32; DB 2; Length 14;  
Best Local Similarity 77.8%; Pred. No. 3.8;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSR 10  
IIII I I I  
Db 4 RRRCVCVR 12

RESULT 11  
US-08-685-589A-194  
Sequence 194, Application US/08685589A  
Patent No. 5916872  
GENERAL INFORMATION:  
APPLICANT: Chang, Conway  
APPLICANT: Gu, Leo  
APPLICANT: Chen, Jie  
TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
TITLE OF INVENTION: SPECTRUM ANTIMICROBIAL ACTIVITY  
NUMBER OF SEQUENCES: 222  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,589A  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-026-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 194:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..14  
OTHER INFORMATION: /product= "Cyclic"  
US-08-685-589A-194

Query Match 65.3%; Score 32; DB 2; Length 14;  
Best Local Similarity 77.8%; Pred. No. 3.8;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSVR 10  
|||||  
DB 4 RRRFCVCVR 12

RESULT 12  
US-08-685-589A-222  
; Sequence 222, Application US/08685589A  
; Patent No. 5916872  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Leo  
; TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
; TITLE OF INVENTION: SPECTRUM ANTIMICROBIAL ACTIVITY  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,589A  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-026-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 222:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..14  
OTHER INFORMATION: /product= "Cyclic"

US-08-685-589A-222  
Query Match 65.3%; Score 32; DB 2; Length 14;  
Best Local Similarity 77.8%; Pred. No. 3.8;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSVR 10  
|||||  
DB 4 RRRFCVCVR 12

RESULT 13  
US-08-685-589A-158  
; Sequence 158, Application US/08685589A  
; Patent No. 5916872  
; GENERAL INFORMATION:

APPLICANT: Chang, Conway  
APPLICANT: Gu, Leo  
APPLICANT: Chen, Jie  
TITLE OF INVENTION: CYCLIC PEPTIDES HAVING BROAD  
TITLE OF INVENTION: SPECTRUM ANTIMICROBIAL ACTIVITY  
NUMBER OF SEQUENCES: 222  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,589A  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-026-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 158:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..14  
OTHER INFORMATION: /product= "Cyclic"

US-08-685-589A-158  
Query Match 63.3%; Score 31; DB 2; Length 14;  
Best Local Similarity 66.7%; Pred. No. 5.8;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSVR 10  
|||||  
DB 4 RRRFCVCVR 12

RESULT 14  
US-08-599-171A-26  
; Sequence 26, Application US/08599171A  
; Patent No. 581473  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Patrick V.  
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:



MEDIUM TYPE: 3.5 INCH DISKETTE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WORD PERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/599,171A  
 FILING DATE: Concurrently  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: HERRON, CHARLES J.  
 REGISTRATION NUMBER: 28,019  
 REFERENCE/DOCKET NUMBER: 331400-38  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 26:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 373 AMINO ACIDS  
 TYPE: AMINO ACID  
 TOPOLOGY: LINEAR  
 MOLECULE TYPE: PROTEIN  
 US-08-599-171A-26

Query Match 63.3%; Score 31; DB 2; Length 373;  
 Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRRRFSVSV 9  
 ||::||| |  
 Db 73 YRKYSVEV 81

RESULT 15  
 US-08-646-590B-26  
 Sequence 26, Application US/08646590B  
 Patent No. 5962283  
 GENERAL INFORMATION:  
 APPLICANT: Warren, Patrick V.  
 APPLICANT: Swanson, Ronald V.  
 TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson, P.C.  
 STREET: 4225 Executive Square, Suite 1400  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: US  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows95  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/646,590B  
 FILING DATE: 08-May-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/599,171  
 FILING DATE: 09-FEB-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US97/01094  
 FILING DATE: 21-January-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Halle, Ph.D., Lisa A.  
 REGISTRATION NUMBER: 38,347  
 REFERENCE/DOCKET NUMBER: 09010/017001

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619/678-5070  
 TELEFAX: 619/678-5099  
 INFORMATION FOR SEQ ID NO: 26:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 373 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 US-08-646-590B-26

Query Match 63.3%; Score 31; DB 2; Length 373;  
 Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRRRFSVSV 9  
 |::||| |  
 Db 73 YRKYSVEV 81

Search completed: February 12, 2002, 12:32:23  
 Job time: 451 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:38 ; Search time 126.85 Seconds  
(without alignments)  
6.005 Million cell updates/sec

Title: US-09-485-571-22  
Perfect score: 49  
Sequence: 1 YRRRFSVSVR 10  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	79.6	113	2 A41830	hypothetical prote
2	39	79.6	151	1 Q2BPP4	hypothetical prote
3	36	73.5	211	2 A46458	human CRI homolog
4	36	73.5	719	2 A83127	probable toxin tra
5	36	73.5	896	2 T27376	hypothetical prote
6	36	73.5	965	2 T38430	hypothetical prote
7	35	71.4	328	2 T00547	hypothetical prote
8	34	69.4	146	2 D82876	hypothetical prote
9	34	69.4	287	2 F82380	transcription regu
10	34	69.4	385	2 S43540	YSD83 protein - ye
11	34	69.4	385	2 S38191	hypothetical prote
12	33	67.3	249	2 B69343	conserved hypotet
13	33	67.3	1188	2 T19552	hypothetical prote
14	32	65.3	146	2 T33956	hypothetical prote
15	32	65.3	167	2 T22726	hypothetical prote
16	32	65.3	325	1 QQBE84	ULF6 protein - hum
17	32	65.3	344	2 T46928	hypothetical prote
18	32	65.3	353	1 D65025	hypothetical prote
19	32	65.3	353	2 D85892	probable permease
20	32	65.3	385	2 H83766	aspartate aminotra
21	32	65.3	391	2 T09058	butyrophilin homol
22	32	65.3	395	2 T24578	hypothetical prote
23	32	65.3	415	2 C84709	hypothetical prote
24	32	65.3	419	2 T33004	hypothetical prote
25	32	65.3	450	2 B83530	probable RNA methy
26	32	65.3	511	1 S54720	glucose-6-phosphat
27	32	65.3	558	2 T15448	hypothetical prote
28	32	65.3	633	2 T17262	hypothetical prote
29	32	65.3	1001	2 T13807	potassium channel

30	32	65.3	1156	2 T23308	hypothetical prote
31	32	65.3	1556	2 S59393	probable membrane
32	32	65.3	2337	2 T40577	probable phosphati
33	31	63.3	107	2 H82729	hypothetical prote
34	31	63.3	124	2 E70754	hypothetical prote
35	31	63.3	199	2 H69429	hypothetical prote
36	31	63.3	282	2 T20402	hypothetical prote
37	31	63.3	283	2 T26682	hypothetical prote
38	31	63.3	303	2 T04681	hypothetical prote
39	31	63.3	333	1 S39659	monooxygenase homo
40	31	63.3	345	2 G85890	lipoprotein-34 lim
41	31	63.3	345	2 D65023	lipoprotein-34 pre
42	31	63.3	373	2 E70338	probable aspartate
43	31	63.3	426	2 E96605	probable peptide c
44	31	63.3	463	2 T04686	hypothetical prote
45	31	63.3	662	2 T23271	hypothetical prote

ALIGNMENTS

RESULT 1

A41830  
hypothetical protein 151 - phage phi-R73  
C:Species: phage phi-R73  
C:Date: 24-Jul-1992 #sequence\_revision 24-Jul-1992 #text\_change 20-Sep-1999  
C:Accession: A41830

R:Sun, J.; Inouye, M.; Inouye, S.  
J. Bacteriol. 173, 4171-4181, 1991  
A:Title: Association of a retroelement with a P4-like cryptic prophage (retroprophage)  
A:Reference number: A42465; MUID:91286203  
A:Accession: A41830  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-113 <SUN>  
A:Cross-references: GB:M64113  
C:Superfamily: satellite phage P4 hypothetical 17.7K protein

Query Match 79.6% Score 39; DB 2; Length 113;  
Best Local Similarity 70.0% Pred. No. 1;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRRRFSVSVR 10  
Db 45 YRRKFSILVR 54  
:::|::|

RESULT 2

Q2BPP4  
hypothetical protein ORF2 - satellite phage P4  
N:Alternate names: hypothetical 17.7K protein; ORF151 protein  
C:Species: satellite phage P4  
A:Note: host Escherichia coli  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 10-Sep-1999  
C:Accession: H23878; JW0026  
R:Lin, C.S.  
Nucleic Acids Res. 12, 8667-8684, 1984  
A:Title: Nucleotide sequence of the essential region of bacteriophage P4.  
A:Reference number: A93544; MUID:85062840  
A:Accession: H23878  
A:Molecule type: DNA  
A:Residues: 1-151 <LIN>  
A:Cross-references: GB:X02534; NID:q15167; PIDN:CAA26380.1; PID:q15174  
R:Halling, C.; Calendar, R.; Christie, G.E.; Dale, E.C.; Deho, G.; Finkel, S.; Fien

M.G.; Ziermann, R.  
Nucleic Acids Res. 18, 1649, 1990  
A:Title: DNA sequence of satellite bacteriophage P4.  
A:Reference number: JW0020; MUID:90221913  
A:Accession: JW0026  
A:Molecule type: DNA  
A:Residues: 1-151 <HAL>  
A:Cross-references: GB:X51522; NID:g450916; PIDN:CAA35900.1; PID:g15160

C:Superfamily: satellite phase P4 hypothetical 17.7k protein

Query Match 79.6%; Score 39; DB 1; Length 151;  
Best Local Similarity 70.0%; Pred. No. 1.4;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRRRFSVSR 10  
|||||: ||  
Db 45 YRRKFSILVR 54

RESULT 3  
A46458

human CR1 homolog CRRP - guinea pig  
C:Species: Cavia porcellus (guinea pig)  
C:Date: 18-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 07-Jul-1995  
C:Accession: A46458  
R:Moore Jr., F.D.  
J. Immunol. 147, 3615-3622, 1991  
A:Title: CRRP: a guinea pig protein, identified by sequence homology to human CR1, which  
A:Reference number: A46458; MUID:92043737  
A:Accession: A46458  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-211 <MOO>  
A>Note: sequence inconsistent with the nucleotide translation  
A>Note: sequence extracted from NCBI backbone (NCBIN:63906, NCBIP:63908)  
C:Superfamily: complement factor H repeat homology  
F:56-112/Domain: complement factor H repeat homology <FH21>  
F:117-174/Domain: complement factor H repeat homology <FH22>

Query Match 73.5%; Score 36; DB 2; Length 211;  
Best Local Similarity 75.0%; Pred. No. 7.5;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRRFSVS 8  
|||||: ||  
Db 91 YRRRFSIT 98

RESULT 4

A83127  
probable toxin transporter PA4143 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83127  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miziochui, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: A83127  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-719 <STO>  
A:Cross-references: GB:AE004831; GB:AE004091; NID:g9950347; PIDN:AAG07530.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4143  
C:Superfamily: hemolysin secretion protein B; ATP-binding cassette homology

Query Match 73.5%; Score 36; DB 2; Length 719;  
Best Local Similarity 77.8%; Pred. No. 24;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 RRRFSVSR 10  
|||||: ||  
Db 48 RRRFSVSLK 56

RESULT 5

T27376  
hypothetical protein Y75B12B.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T27376  
R:White, S.

submitted to the EMBL Data Library, October 1998  
A:Reference number: Z20360

A:Accession: T27376

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-896 <WIL>

A:Cross-references: EMBL:AL032663; PIDN:CAA21765.1; GSPDB:GN00023; CESP:Y75B12B.6

A:Experimental source: clone Y75B12B

C:Genetics:

A:Gene: CESP:Y75B12B.6

A:Map position: 5

A:Introns: 21/3; 54/3; 115/3; 163/2; 189/3; 223/2; 273/1; 325/1; 642/2; 674/3; 783/1;  
C:Superfamily: squid 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase; 1-ph  
osphodiesterase domain Y homology  
F:306-467/Domain: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X

Query Match 73.5%; Score 36; DB 2; Length 896;  
Best Local Similarity 70.0%; Pred. No. 30;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRRRFSVSR 10  
|||||: ||  
Db 689 FRRSSVSIR 698

RESULT 6

T38430  
hypothetical protein SPAC26H5.11 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T38430  
R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z21748  
A:Accession: T38430  
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-965 <OLI>

A:Cross-references: EMBL:Z99126; PIDN:CAB16196.1; GSPDB:GN00066; SPDB:SPAC26H5.11

A:Experimental source: strain 972h; cosmid c26H5

C:Genetics:

A:Gene: SPDB:SPAC26H5.11

A:Map position: 1

A:Introns: 938/2

Query Match 73.5%; Score 36; DB 2; Length 965;  
Best Local Similarity 50.0%; Pred. No. 32;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRRFSVSR 10  
|||||: ||  
Db 847 YRKYSISLK 856

RESULT 7

T00547  
hypothetical protein At2g39370 [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F12L6.3  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 16-Feb-2001  
C:Accession: T00547; E84816  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.;  
submitted to the EMBL Data Library, July 1998



Wed Feb 13 07:52:03 2002

us-09-485-571-22.rpr

QY 1 YRRFVS SV 9  
:||||:|  
Db 1110 FRRFGLSV 1118

RESULT 14  
T33956  
hypothetical protein C13B7.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T33956  
R:Smith, A.; Wamsley, P.; Hawkins, M.  
A:Description: The sequence of C. elegans cosmid C13B7.  
A:Reference number: Z21445  
A:Accession: T33956  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-146 <SMI>  
A:Cross-references: EMBL:AF125954; PIDN:AD14706.1; GSPDB:GN00023; CBSP:C13B7.4  
A:Experimental source: strain Bristol N2; clone C13B7  
C:Genetics:  
A:Gene: CBSP:C13B7.4  
A:Map position: 5  
A:Introns: 64/2; 96/3

Query Match 69.4%; Score 34; DB 2; Length 385;  
Best Local Similarity 66.7%; Pred. No. 33;  
Matches 2; Conservative 1; Indels 0; Gaps 0;

QY 1 YRRFVS SV 9  
:||||:|  
Db 130 YRRFTVVI 138

RESULT 12  
B69343  
conserved hypothetical protein AF0746 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: B69343  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artisch, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A:Reference number: A69250; MUID:98049343  
A:Accession: B69343  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-249 <KLE>  
A:Cross-references: GB:AE001053; GB:AE000782; NID:g2689376; PIDN:AAB90493.1; PID:g264986  
C:Superfamily: conserved hypothetical protein MJ0144

Query Match 67.3%; Score 33; DB 2; Length 249;  
Best Local Similarity 66.7%; Pred. No. 35;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 RRRFVS SV 10  
:|:|:|:|  
Db 71 RRKFAVNR 79

RESULT 13  
T19552  
hypothetical protein C29E6.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T19552  
R:Dobson, R.  
A:Submitted to the EMBL Data Library, May 1996  
A:Reference number: Z19141  
A:Accession: T19552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1188 <WIL>  
A:Cross-references: EMBL:Z72504; PIDN:CAA96603.1; GSPDB:GN00022; CBSP:C29E6.2  
A:Experimental source: clone C29E6  
C:Genetics:  
A:Gene: CBSP:C29E6.2  
A:Map position: 4  
A:Introns: 147/2; 473/3; 500/1; 529/1; 584/2; 688/2; 839/3; 975/3; 1132/3

Query Match 67.3%; Score 33; DB 2; Length 1188;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRRFVS SV 9  
:||||:|  
Db 1110 FRRFGLSV 1118

RESULT 14  
T33956  
hypothetical protein C13B7.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T33956  
R:Smith, A.; Wamsley, P.; Hawkins, M.  
A:Description: The sequence of C. elegans cosmid C13B7.  
A:Reference number: Z21445  
A:Accession: T33956  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-146 <SMI>  
A:Cross-references: EMBL:AF125954; PIDN:AD14706.1; GSPDB:GN00023; CBSP:C13B7.4  
A:Experimental source: strain Bristol N2; clone C13B7  
C:Genetics:  
A:Gene: CBSP:C13B7.4  
A:Map position: 5  
A:Introns: 64/2; 96/3

Query Match 65.3%; Score 32; DB 2; Length 146;  
Best Local Similarity 70.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRRFVS SV 10  
:||||:|  
Db 112 YRRRLSSVR 121

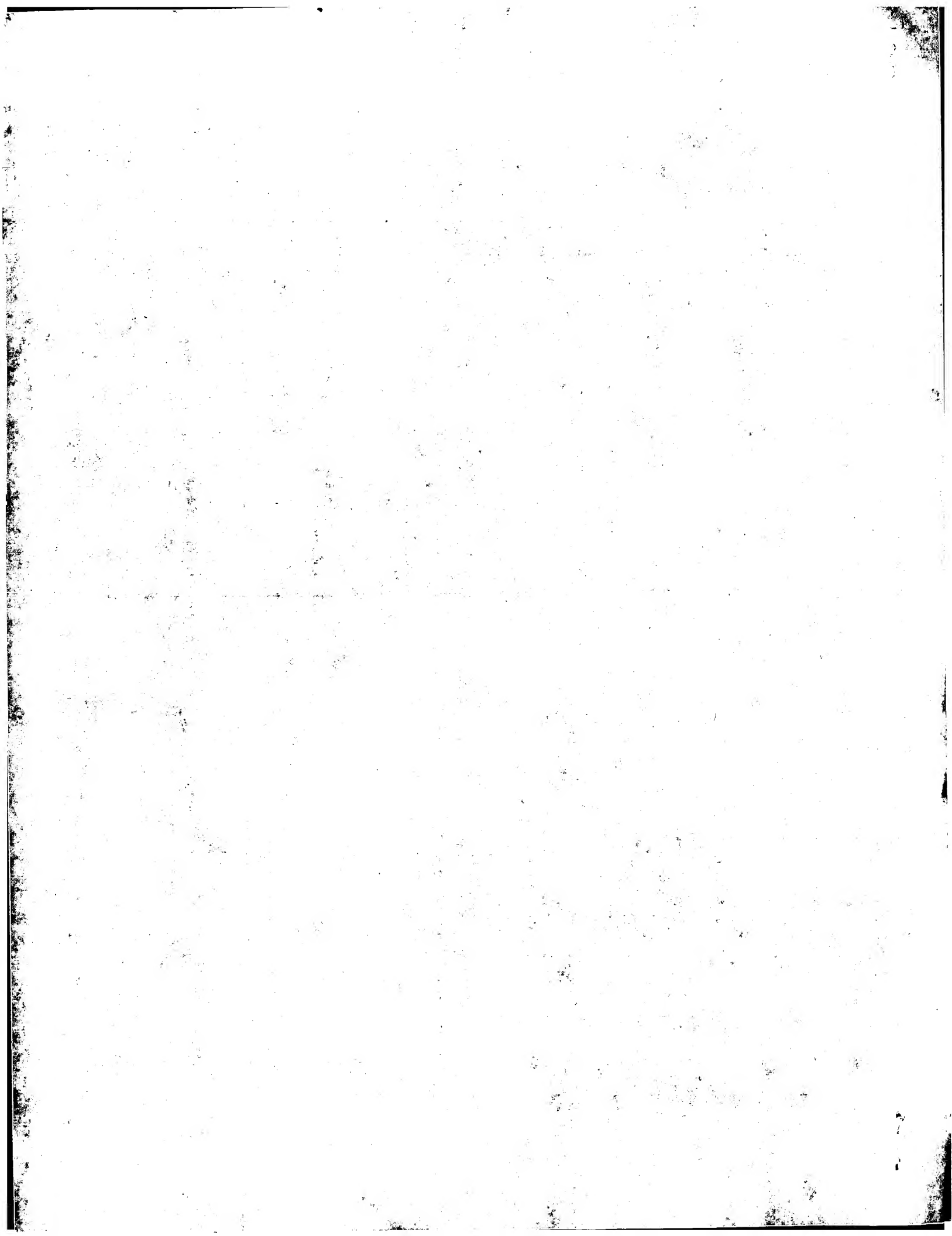
RESULT 15  
T22726  
hypothetical protein F55C9.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T22726  
R: Basham, V.  
A:Submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19605  
A:Accession: T22726  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-167 <WIL>  
A:Cross-references: EMBL:Z81549; PIDN:CAB04473.1; GSPDB:GN00023; CBSP:F55C9.3  
A:Experimental source: clone F55C9  
C:Genetics:  
A:Gene: CBSP:F55C9.3  
A:Map position: 5  
A:Introns: 32/2; 75/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein F55C9.3

Query Match 65.3%; Score 32; DB 2; Length 167;  
Best Local Similarity 50.0%; Pred. No. 37;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRRFVS SV 10  
:||||:|:|  
Db 93 FRRYINVELR 102

Search completed: February 12, 2002, 12:34:39  
Job time: 557 sec

---





Result No.	Query	Score	Match	Length	DB	ID	Description
1	39	79.6	151	1	Y17K_BPP4	P05464	bacterioph
2	34	69.4	385	1	YHH7_SACDO	P41907	saccharomyc
3	34	69.4	385	1	YHH7_YEAST	P32792	saccharomyc
4	33	67.3	93	1	AAT7_METEX	P52069	methylobact
5	33	67.3	604	1	PGH2_MUSVY	O62725	mustela vis
6	32	67.3	604	1	UL76_HCMVA	P16725	human cytom
7	32	65.3	353	1	PERM_ECOLI	P77406	escherichia
8	32	65.3	510	1	GSPD_ASPNG	P48826	aspergillus
9	32	65.3	511	1	GSPD_EMENI	P41764	emeritella
10	32	65.3	1001	1	ORK1_DROME	P494526	drosophila
11	31	63.3	124	1	YC69_MYCTY	Q11050	mycobacteri
12	31	63.3	333	1	YNCH_BACSU	P39606	bacillus su
13	31	63.3	344	1	NLPB_ECOLI	P21167	escherichia
14	30	61.2	74	1	LZMU_ADEMI	O10443	mouse adeno
15	30	61.2	232	1	THIF_PEA	P29450	pisum sativ
16	30	61.2	181	1	NARW_ECOLI	P19317	escherichia
17	30	61.2	327	1	YJBN_HAETN	P44794	haemophilus
18	30	61.2	390	1	VASS_BPGA	P07394	bacterioph
19	30	61.2	484	1	YLS5_CAEEL	P34390	caenorhabdi
20	30	61.2	489	1	YENS_YEAST	P39970	saccharomyc
21	30	61.2	495	1	YDQ7_SCHPO	O14200	schizosacch
22	30	61.2	543	1	P69_MYCGE	P47533	mycoplasma
23	30	61.2	578	1	ACEK_ECOLI	P11071	escherichia
24	30	61.2	583	1	ACEK_SALTU	P51067	salmonella
25	30	61.2	604	1	PGH2_MOUSE	Q05769	m prostagla
26	30	61.2	604	1	PGH2_RAT	P35355	rattus norv
27	30	61.2	638	1	LIK2_HUMAN	P33671	homo sapien
28	30	61.2	638	1	LIK2_RAT	P53670	rattus norv
29	30	61.2	942	1	HEX_ADEG1	P42671	avian adeno
30	30	61.2	1160	1	CIDB_BACTU	Q45747	bacillus th
31	30	61.2	1235	1	KPB2_HUMAN	P46019	homo sapien
32	30	61.2	1235	1	KPB2_RABIT	P46018	oryctolagus
33	29	59.2	59	1	HSPL_PHACI	P79990	phascoglarct



```
Query Match 67.3%; Score 33; DB 1; Length 93;
Best Local Similarity 66.7%; Pred. No. 3.2;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRRRFVSU 9
Db 81 YORRFGVSL 89

RESULT 5
PGH2_MUSVI
ID PGH2_MUSVI STANDARD; PRT; 604 AA.
AC O62725;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROSTAGLANDIN G/H SYNTHASE 2 PRECURSOR (EC 1.14.99.1) (CYCLOOXYGENASE
DE -2) (COX-2) (PROSTAGLANDIN-ENDOPEROXIDE SYNTHASE 2) (PROSTAGLANDIN H2
DE SYNTHASE 2) (PGH SYNTHASE 2) (PGHS-2) (PHS II).
GN PTGS2 OR COX2.
OS Mustela vison (American mink).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustela.
OX NCBI_TaxID=9667;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=98344842; PubMed=9681517;
RA Song J.H., Sirols J., Houde A., Murphy B.D.;
RT "Cloning, developmental expression, and immunohistochemistry of
RT cyclooxygenase 2 in the endometrium during embryo implantation and
RT gestation in the mink (Mustela vison).";
RL Endocrinology 139:3629-3636(1998).
CC -1- FUNCTION: MAY HAVE A ROLE AS A MAJOR MEDIATOR OF INFLAMMATION
CC AND/OR A ROLE FOR PROSTANOID SIGNALING IN ACTIVITY-DEPENDENT
CC PLASTICITY.
CC -1- CATALYTIC ACTIVITY: ARACHIDONATE + AH(2) + 2 O(2) = PROSTAGLANDIN
CC H2 + A + H(2)O.
CC -1- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND
CC THROMBOXANES.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MICROSOMAL MEMBRANE.
CC -1- MISCELLANEOUS: THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A
CC PEROXIDASE.
CC -1- MISCELLANEOUS: THIS ENZYME IS THE TARGET OF NONSTEROIDAL
CC ANTI-INFLAMMATORY DRUGS SUCH AS ASPIRIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AF047841; AAC05637.1; -.
DR InterPro: IPR002007; Anim_peroxidse.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001536; Peroxidse_3.
DR PRINTS: PR00457; ANPEROXIDASE.
DR SMART: SM00181; EGF: 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; FALSE_NEG.
KW Oxidoreductase; Dioxygenase; Peroxidase; Glycoprotein; Acetylation;
KW Prostaglandin biosynthesis; Heme; Iron; Signal; Membrane.
FT SIGNAL 1 17 BY SIMILARITY
FT CHAIN 18 604 PROSTAGLANDIN G/H SYNTHASE 2.
FT DOMAIN 18 55 EGF-LIKE.
FT ACT_SITE 193 193 DISTAL HISTIDINE (BY SIMILARITY).
FT ACT_SITE 371 371 CYCLOOXYGENASE (BY SIMILARITY).

Query Match 65.3%; Score 32; DB 1; Length 325;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRRRFVSU 10
Db 63 YRRRLFTEVR 72

RESULT 6
UL76_HCMVA
ID UL76_HCMVA STANDARD; PRT; 325 AA.
AC P16723;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE PROTEIN UL76.
GN UL76.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohnl R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169.";
RL Curr Top Microbiol Immunol. 154:125-169(1990).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL24,
CC EHV-1 37, EBV BRF1, HCMV UL76, ILTV ORF3, AND VZV 35.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: X17403; CAA35391.1; -.
DR PIR: S09839; QBEBA4.
DR InterPro: IPR002580; Herpes_UL24.
DR Pfam: PF01646; Herpes_UL24; 1.
SQ SEQUENCE 325 AA; 36069 MW; 5BCD72EC8E8F9BF9 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 325;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YRRRFVSU 10
Db 63 YRRRLFTEVR 72
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
CC EMBL; AE000335; AAC75546.1;
CC DR EMBL; D90878; BAA16382.1;
CC DR EMBL; AE005479; AAG57603.1;
CC DR EMBL; AP002561; BAB36778.1;
CC DR EMBL; M63654; AAB88575.1; ALT_SEQ.
CC EcoGene; EGI4221; perm.
CC DR InterPro; IPR002549; UPF0118.
CC DR Pfam; PF01594; UPF0118.1.
CC KW Transmembrane; Transport; Complete proteome.
CC FT TRANSMEM 19 39 POTENTIAL.
CC FT TRANSMEM 72 92 POTENTIAL.
CC FT TRANSMEM 156 176 POTENTIAL.
CC FT TRANSMEM 217 237 POTENTIAL.
CC FT TRANSMEM 240 260 POTENTIAL.
CC FT TRANSMEM 281 301 POTENTIAL.
CC FT TRANSMEM 310 330 POTENTIAL.
CC SQ SEQUENCE 353 AA; 39194 MW; EICB8EB1E242FC3B CRC64;

Query Match 65.3%; Score 32; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRRFS 6
DB 9 YRRRFS 14

RESULT 8
G6PD_ASPNG STANDARD; PRT; 510 AA.
ID G6PD_ASPNG
AC P48826;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).
GN GSDA OR G6PDH.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-NRRL 3;
RA Thamm A.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-CBS 120.49 / N400;
RX MEDLINE=95272533; PubMed=7753033;
RA van den Broek P., Goosen T., Wennekes B., van den Broek H.;
RT "Isolation and characterization of the glucose-6-phosphate
RT dehydrogenase encoding gene (gsdA) from Aspergillus niger.";
RL Mol. Gen. Genet. 247:229-239(1995).
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) -> D-GLUCONO-
CC DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

RESULT 7
PERM_ECOLI STANDARD; PRT; 353 AA.
AC P77406; P71230;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE PERMEASE PERM.
GN PERM OR B2493 OR Z3755 OR ECS3355.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubraman S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RA "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [5]
RN SEQUENCE OF 279-353 FROM N.A.
RC STRAIN-K12;
RA Andrews S.C., McClay J., Ambler A., Quail M., Berks B.C., Guest J.R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0118 (PERM) FAMILY. STRONG, TO
CC H.INFLUENZAE H10237/38.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; X87942; CAA61194.1; -.
DR EMBL; X77829; CAA54840.1; -.
DR HSSP; P11411; IDPG.
DR InterPro; IPR001282; G6PD.
DR Pfam; PF00479; G6PD; 1.
DR PRINTS; PR00079; G6PDHGRNASE.
DR ProDom; PD001129; G6PD; 1.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Glucose metabolism.
FT ACT_SITE 198 198 BY SIMILARITY.
FT CONFLICT 135 135 L -> F (IN REF. 2).
FT CONFLICT 508 510 NRL -> TVCK (IN REF. 2).
SQ SEQUENCE 510 AA; 58950 MW; FDDF3F5025483AF6 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 510;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRRFS 6
DB 413 YRRRFS 418

RESULT 9
ID G6PD_EMENI STANDARD; PRT; 511 AA.
AC P41764; Q92408;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).
GN GSDA OR G6PD.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RA Schaap P.J., Muller Y., Visser J.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA van den Broek P., Goosen T., Wennekes B., van den Broek H.;
RC Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RL Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC -|- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCONO-
CC DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC -|- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC -|- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X84001; CAA58825.1; -.
DR EMBL; X77830; CAA54841.1; -.
DR HSSP; P11411; 2DPG.
DR InterPro; IPR001282; G6PD.
DR Pfam; PF00479; G6PD; 1.
DR PRINTS; PR00079; G6PDHGRNASE.
DR ProDom; PD001129; G6PD; 1.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Glucose metabolism.
FT ACT_SITE 198 198 BY SIMILARITY.
FT CONFLICT 15 20 MISSING (IN REF. 1).
FT CONFLICT 76 85 RSYIKPTKE -> DTLRPRQRK (IN REF. 1).

```

```

FT CONFLICT 352 352 L -> LP (IN REF. 1).
SQ SEQUENCE 511 AA; 58977 MW; 669C15B72878A475 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 511;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRRFS 6
DB 413 YRRRFS 418

RESULT 10
ID ORK1_DROME STANDARD; PRT; 1001 AA.
AC Q94526;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OPEN RECTIFIER POTASSIUM CHANNEL PROTEIN 1 (TWO PORE DOMAIN POTASSIUM
DE CHANNEL ORK1).
DE CHANNEL ORK1.
GN ORK1 OR CGI615.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Larva;
RX MEDLINE=97075152; PubMed=8917578;
RA Goldstein S.A.N., Price L.A., Rosenthal D.N., Pausch M.H.;
RT "ORK1, a potassium-selective leak channel with two pore domains
RT cloned from Drosophila melanogaster by expression in Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13256-13261(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel E.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

```

Wed Feb 13 07:52:03 2002

us-09-485-571-22.rsp

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Zheng L.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2183-2195(2000).  
 CC -!- FUNCTION: BACKGROUND POTASSIUM CHANNEL. RECTIFICATION IS DEPENDENT  
 CC ON EXTERNAL POTASSIUM CONCENTRATION. ACTS AS AN OUTWARDLY  
 CC RECTIFYING CHANNEL BUT AS EXTERNAL POTASSIUM LEVELS INCREASE, THIS  
 CC IS REVERSED.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -!- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN ADULT, STRONGEST  
 CC EXPRESSION IN MUSCLE, BRAIN AND OVARY. ALSO PRESENT AT LOW LEVELS  
 CC IN LARVA AND EMBRYO.  
 CC -!- MISCELLANEOUS: INHIBITED BY BARIUM.  
 CC -!- SIMILARITY: BELONGS TO THE TWO PORE DOMAIN FAMILY OF POTASSIUM  
 CC CHANNELS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U55321; AAC69250.1; -;  
 CC EMBL: AF003484; AAF47972.1; -;  
 CC FlyBase: FBgn0017561; Ork1.  
 CC InterPro: IPR003280; 2porek\_channel.  
 CC InterPro: IPR001622; Channel\_pore\_k.  
 CC InterPro: IPR000099; TWIK\_channel.  
 CC Pfam: PF02034; TWIK\_channel. 1.  
 CC PRINTS: PR01333; 2POREKCHANNEL.  
 CC Ionic channel; Transmembrane; Ion transport; Potassium transport;  
 KW Glycoprotein. 1 6 CYTOPLASMIC (POTENTIAL).  
 KW DOMAIN 1 6 POTENTIAL.  
 KW TRANSNEM 7 27 PORE-FORMING (POTENTIAL).  
 KW DOMAIN 95 111 POTENTIAL.  
 KW TRANSNEM 120 140 POTENTIAL.  
 KW DOMAIN 141 170 CYTOPLASMIC (POTENTIAL).  
 KW TRANSNEM 171 191 POTENTIAL.  
 KW DOMAIN 208 224 PORE-FORMING (POTENTIAL).  
 KW TRANSNEM 244 264 POTENTIAL.  
 KW DOMAIN 265 1001 CYTOPLASMIC (POTENTIAL).  
 KW CARBOHYD 58 58 N-LINKED (GLCNAC...). (POTENTIAL).  
 SQ SEQUENCE 1001 AA; 109289 MW; 09AE1A3669072E07 CRC64;  
 Query Match 65.3%; Score 32; DB 1; Length 1001;  
 Best Local Similarity 85.7%; Pred. No. 72;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YRRFESV 7  
 DB 875 FRRFESV 881  
 :|||||  
 RESULT 11  
 YG69\_MYCTU  
 ID YC69\_MYCTU STANDARD; PRT; 124 AA.  
 AC Q11050;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN RV1269C PRECURSOR.  
 GN RV1269C OR MT1307 OR MYC50.13.  
 OS *Mycobacterium tuberculosis*.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; *Mycobacterium*.  
 OC Actinomycetales; Corynebacterineae; *Mycobacteriaceae*; *Mycobacterium*.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RN SEQUENCE FROM N.A.

RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischnmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN IDENTIFICATION BY MASS SPECTROMETRY.  
 RC STRAIN=H37RV;  
 RA Mattow J., Jungblut P.R., Mueller E.-C., Kaufmann S.H.E.;  
 RT "Identification of acidic, low molecular mass proteins of  
 RT *Mycobacterium tuberculosis* strain H37RV by matrix-assisted laser  
 RT desorption/ionization and electrospray ionization mass spectrometry.";  
 RL Proteomics 1:494-507(2001).  
 CC -!- SIMILARITY: TO *M.TUBERCULOSIS* RV1813C.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z77137; CAB00910.1; -;  
 CC EMBL: AE007005; AAK45567.1; ALT\_INIT.  
 CC TIGR: MT1307; -;  
 CC TuberculList: RV1269c; -;  
 CC Signal: Complete proteome.  
 CC SIGNAL 1 35 POTENTIAL.  
 CC CHAIN 36 124 PROTEIN RV1269C.  
 CC SEQUENCE 124 AA; 12550 MW; 1BCECB3EFAB4FC0C CRC64;  
 Query Match 63.3%; Score 31; DB 1; Length 124;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 RRRFSVS 9  
 DB 8 RRRFAV 15  
 :|||||  
 RESULT 12  
 YWCH\_BACSU  
 ID YWCH\_BACSU STANDARD; PRT; 333 AA.  
 AC P39606;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL 36.6 KDA PROTEIN IN QOXD-VPR INTERGENIC REGION.  
 GN YWCH OR IPA-44D.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; *Bacillus*; *Clostridium* group;  
 OC *Bacillus/Staphylococcus* group; *Bacillus*.

```

OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.A., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
CC -!- FUNCTION: NON-ESSENTIAL LIPOPROTEIN.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR.
CC -!- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73124; CAA51600.1; -;
DR PIR: S39699; S39699.
DR Subtilist; BG10590; ywch.
DR InterPro: IPR002103; Bac.Luciferase.
DR Pfam: PF00296; bac_luciferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 333 AA; 36557 MW; B593613BB8FD8BC5 CRC64;

Query Match 63.3%; Score 31; DB 1; Length 333;
Best Local Similarity 56.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRRRFSVSV 9
DB 205 YKRFKPSV 213
I:|||||
I:|||||

RESULT 13
NLPB_ECOLI
ID NLPB_ECOLI STANDARD; PRT; 344 AA.
AC P21167; P76564;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIPOPROTEIN-34 PRECURSOR.
GN NLPB OR DAPX OR B2477.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91358331; PubMed=1885529;
RA Bouvier J., Pugsley A.P., Stragier P.;
RT "A gene for a new lipoprotein in the dapa-purC interval of the
RT Escherichia coli chromosome.";
RL J. Bacteriol. 173:5523-5531(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91008982; PubMed=2120198;
RA Tiedemann A.A., Denarini D.J., Parker J., Smith J.M.;
RT "DNA sequence of the purC gene encoding 5'-phosphoribosyl-5-
RT aminimidazole-4-N-succinocarboxamide synthetase and organization of
RT the dapa-purC region of Escherichia coli K-12.";
RL J. Bacteriol. 172:6035-6041(1990).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.A., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
CC -!- FUNCTION: NON-ESSENTIAL LIPOPROTEIN.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR.
CC -!- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57402; CAA40661.1; -;
DR PIR: S25426; S25426.
DR EcoGene; EG10658; nlpB
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal; Complete proteome.
FT SIGNAL 1 24
FT CHAIN 25 344 LIPOPROTEIN-34.
FT LIPID 25 25 N-ACYL DIGLYCERIDE.
SQ SEQUENCE 344 AA; 36842 MW; 49991F277D9D923C CRC64;

Query Match 63.3%; Score 31; DB 1; Length 344;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRRRFSVSVR 10
DB 157 YRGYQISVK 166
I:|||||
I:|||||

RESULT 14
L2MU_ADEM1
ID L2MU_ADEM1 STANDARD; PRT; 74 AA.
AC O10443;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LATE L2 MU CORE PROTEIN PRECURSOR (PMU) (PROTEIN X).
GN PX.
OS Mouse adenovirus type 1 (MAV-1).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.

```

Wed Feb 13 07:52:03 2002

```

NCBI_TaxID-10530;
[1]
SEQUENCE FROM N.A.
RA Meissner J.D., Hirsch G.N., Larue E.A., Fulcher R.A., Spindler K.R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE ROLE OF THE PRECURSOR MIGHT BE TO CONDENSE THE
CC VIRAL PROCHROMATIN FOR ENCAPSIDATION BY VIRTUE OF THE TWO
CC BASIC DOMAINS (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X63537; CAA45098.1; -
CC EMBL; U35830; AAC49357.1; -
CC PIR; S20929; S20929.
CC HSP; P10599; IAUU.
CC InterPro; IPR000063; ThioRed.
CC Pfam; PF00085; thioRed; 1.
CC PROSITE; PS00194; THIOREDOXIN; 1.
CC Redox-active center; Electron transport; chloroplast; Transit peptide.
CC TRANSIT 1 69 CHLOROPLAST (POTENTIAL).
CC CHAIN 70 182 THIOREDOXIN F-TYPE.
CC DISULFID 106 109 REDOX-ACTIVE (BY SIMILARITY).
CC SEQUENCE 182 AA; 19775 MW; 158FC352CB9E0FF1 CRC64;

Query Match 61.2%; Score 30; DB 1; Length 182;
Best Local Similarity 77.8%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RRRFSVSVR 10
DB 54 RRSISVSVR 62

Search completed: February 12, 2002, 12:39:51
Job time: 804 sec

```

```

NCBI_TaxID-10530;
[1]
SEQUENCE FROM N.A.
RA Meissner J.D., Hirsch G.N., Larue E.A., Fulcher R.A., Spindler K.R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE ROLE OF THE PRECURSOR MIGHT BE TO CONDENSE THE
CC VIRAL PROCHROMATIN FOR ENCAPSIDATION BY VIRTUE OF THE TWO
CC BASIC DOMAINS (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U95843; AAB53758.1; -
CC Core protein; DNA-binding; Late protein.
CC PROPEP 1 27 POTENTIAL.
CC CHAIN 28 43 LATE L2 MU CORE PROTEIN.
CC PROPEP 44 74 POTENTIAL.
CC SITE 27 28 CLEAVAGE (BY ADENOVIRUS PROTEASE)
CC SITE 43 44 CLEAVAGE (BY ADENOVIRUS PROTEASE)
CC SITE 43 44 CLEAVAGE (BY ADENOVIRUS PROTEASE)
CC SEQUENCE 74 AA; 8316 MW; 647AC6A52D9670AB CRC64;

Query Match 61.2%; Score 30; DB 1; Length 74;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

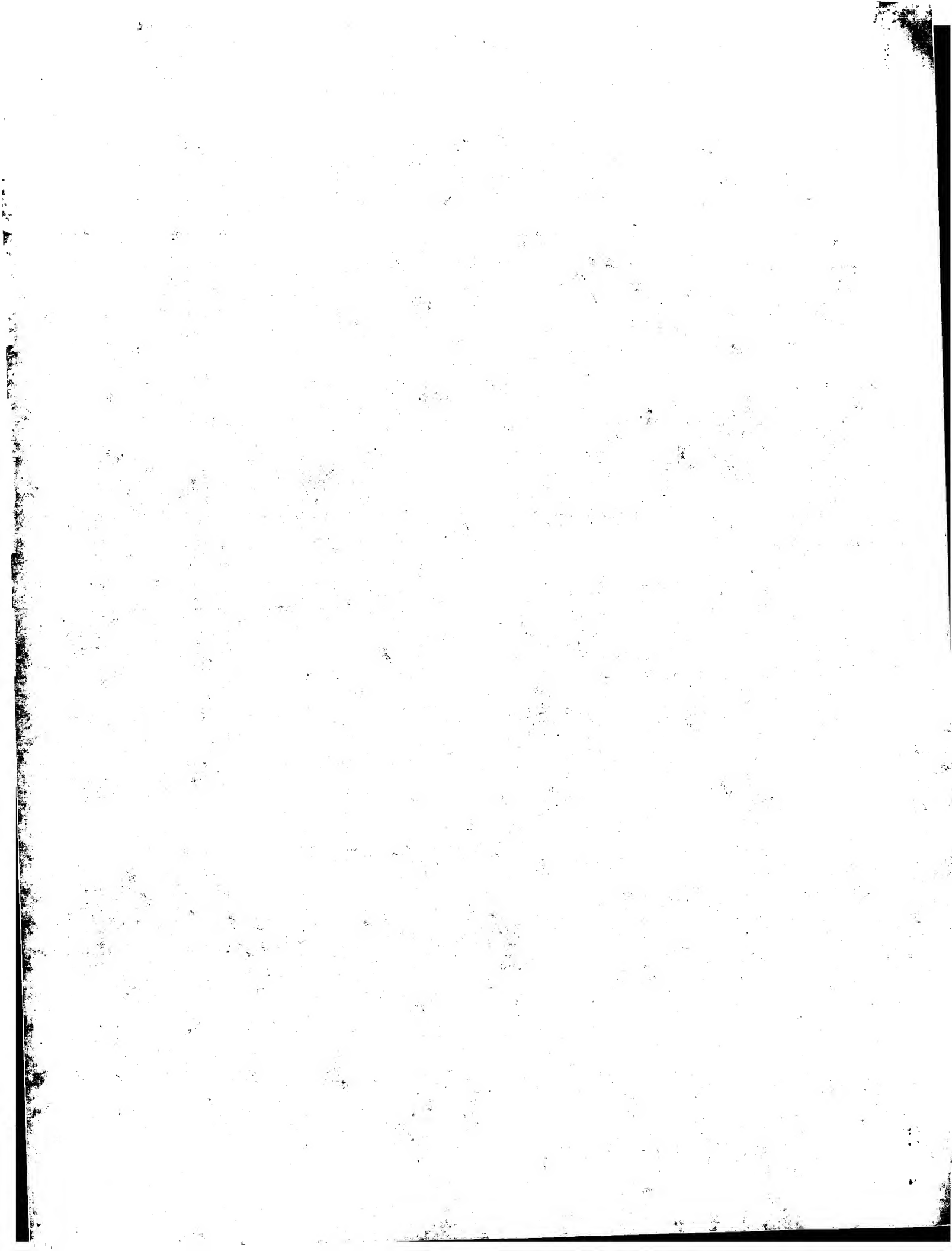
QY 1 YRRFVSFVSVR 10
DB 8 YRFRFPVALR 17

RESULT 15
THIF-PEA STANDARD; PRT; 182 AA.
ID THIF-PEA
AC P29450;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE THIOREDOXIN F-TYPE, CHLOROPLAST PRECURSOR (TRX-F).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=92256804; PubMed=1581563;
RA Lepiniec L., Hodges M., Gadal P., Cretin C.;
RT "Isolation, characterization and nucleotide sequence of a full-length
RT pea cDNA encoding thioredoxin-f."
RL Plant Mol. Biol. 18:1023-1025(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96215867; PubMed=8642611;
RA Sahrway M., Hecht V., Lopez Jaramillo J., Chueca A., Chartier Y.,
RA Meyer Y.;
RT "Intron position as an evolutionary marker of thioredoxins and
RT thioredoxin domains."
RL J. Mol. Evol. 42:422-431(1996).
CC -!- FUNCTION: PARTICIPATES IN VARIOUS REDOX REACTIONS THROUGH THE
CC REVERSIBLE OXIDATION OF THE ACTIVE CENTER DITHIOL, TO A DISULFIDE.
CC THE F FORM IS KNOWN TO ACTIVATE A NUMBER OF ENZYMES OF THE
CC PHOTOSYNTHETIC CARBON CYCLE.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. PLANT F-TYPE.

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:39 ; Search time 232.64 Seconds  
(without alignments)  
6.287 Million cell updates/sec

Title: US-09-485-571-22  
Perfect score: 49  
Sequence: 1 YRRRFSVSVR 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	73.5	180	11 Q60400	Q60400 cavia porce
2	36	73.5	719	2 Q9HWN8	Q9HWN8 pseudomonas
3	36	73.5	896	5 Q9XWB7	Q9XWB7 caenorhabdi
4	36	73.5	965	3 O13992	O13992 schizosacch
5	36	73.5	1703	10 Q9SR02	Q9SR02 arabidopsis
6	35	71.4	284	10 Q9FYD4	Q9FYD4 arabidopsis
7	35	71.4	328	10 O80624	O80624 arabidopsis
8	34	69.4	146	2 Q9PPS7	Q9PPS7 ureaplasma
9	34	69.4	287	2 Q9KKM9	Q9KKM9 vibrio chol
10	34	69.4	437	2 Q9CJX3	Q9CJX3 pasteurella
11	34	69.4	1023	2 Q9X6T5	Q9X6T5 bifidobacte
12	33	67.3	125	2 Q9L601	Q9L601 salmonella
13	33	67.3	249	1 O29512	O29512 archaeglob
14	33	67.3	252	10 Q9XGP7	Q9XGP7 oryza sativ
15	33	67.3	513	12 Q9IBW2	Q9IBW2 turkey herp
16	33	67.3	513	12 Q9E6Q4	Q9E6Q4 turkey herp
17	33	67.3	532	13 O12989	O12989 brachydanio
18	33	67.3	589	2 Q9A718	Q9A718 caulobacter
19	33	67.3	651	2 Q9KWR2	Q9KWR2 thermoactin

20	33	67.3	1188	5 Q18297	Q18297 caenorhabdi
21	33	67.3	1328	13 P79754	P79754 fuqu rubrip
22	32	65.3	146	5 Q9AUU6	Q9AUU6 caenorhabdi
23	32	65.3	153	2 Q9L266	Q9L266 streptomyc
24	32	65.3	167	5 Q9XU19	Q9XU19 caenorhabdi
25	32	65.3	294	5 O45140	O45140 caenorhabdi
26	32	65.3	301	6 Q9GM27	Q9GM27 macaca fasc
27	32	65.3	323	5 Q9N4V6	Q9N4V6 caenorhabdi
28	32	65.3	344	4 Q9NSQ5	Q9NSQ5 homo sapien
29	32	65.3	385	2 Q9KEB7	Q9KEB7 bacillus ha
30	32	65.3	391	11 O35441	O35441 mus musculu
31	32	65.3	395	5 Q22251	Q22251 caenorhabdi
32	32	65.3	396	4 Q15044	Q15044 homo sapien
33	32	65.3	415	10 O04344	O04344 arabidopsis
34	32	65.3	419	5 O61519	O61519 caenorhabdi
35	32	65.3	450	2 Q9I525	Q9I525 pseudomonas
36	32	65.3	482	4 Q9UKL0	Q9UKL0 homo sapien
37	32	65.3	558	5 Q17797	Q17797 caenorhabdi
38	32	65.3	593	10 Q9M6N8	Q9M6N8 arabidopsis
39	32	65.3	633	4 Q9UFS7	Q9UFS7 homo sapien
40	32	65.3	762	5 Q9VXQ6	Q9VXQ6 drosophila
41	32	65.3	772	4 Q9H912	Q9H912 homo sapien
42	32	65.3	772	4 Q9BWH6	Q9BWH6 homo sapien
43	32	65.3	1156	5 Q21214	Q21214 caenorhabdi
44	32	65.3	1337	4 Q9P2E4	Q9P2E4 homo sapien
45	32	65.3	1556	3 Q06554	Q06554 saccharomyc

ALIGNMENTS

RESULT 1					
Q60400	ID	Q60400	PRELIMINARY;	PRT;	180 AA.
AC	Q60400				
DT	01-NOV-1996	(TrEMBLrel. 01, Created)			
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)			
DE	CRPP.				
OS	Cavia porcellus (Guinea pig).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.				
OX	NCBI_TaxID=10141;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=92043737; PubMed=1834743;				
RX	Moore F.D. Jr.;				
RT	"CRPP: a guinea pig protein, identified by sequence homology to human				
RT	CR1, which contains two short consensus repeat motifs and appears not				
RT	to be transmembrane or secreted.";				
RL	J. Immunol. 147:3615-3622(1991).				
DR	EMBL; N77760; AAA37034.1; ..				
DR	InterPro; IPR000436; Sushi_SCR_CCP.				
DR	Pfam; PF00084; sushi; 2.				
DR	SMART; SM00032; CCP; 2.				
SQ	SEQUENCE 180 AA; 20370 MW; 55A208CA898BA5B0 CRC64;				
Query Match 73.5%; Score 36; DB 11; Length 180;					
Best Local Similarity 75.0%; Pred. No. 12;					
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;					
QY	1 YRRRFSVS 8				
	::				
Db	60 YRRRFSIT 67				
RESULT 2					
Q9HWN8					
ID	Q9HWN8	PRELIMINARY;	PRT;	719 AA.	
AC	Q9HWN8				
DT	01-MAR-2001	(TrEMBLrel. 16, Created)			
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)			



```

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE F7018.23 PROTEIN.
GN F7018.23.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F7018 genomic sequence.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC011437; AAF04900.1; -
SQ SEQUENCE 1703 AA; 185507 MW; 5C68D7E02FB66A39 CRC64;

Query Match 73.5%; Score 36; DB 10; Length 1703;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRRFVSVR 10
Db 1234 YKRFAVDMR 1243
|||||:|

RESULT 6
Q9FYD4
ID Q9FYD4 PRELIMINARY; PRT; 284 AA.
AC Q9FYD4
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE REC-LIKE PROTEIN.
GN F21E1.190.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Mewes H.W., Rudd S., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL391716; CAC05509.1; -
DR InterPro: IPR001594; Znf-DHHC.
DR Pfam: PF01529; zf-DHHC; 1.
DR ProDom: PD03041; Znf-DHHC; 1.
SQ SEQUENCE 284 AA; 32202 MW; 60F8A52CD8F171A4 CRC64;

Query Match 71.4%; Score 35; DB 10; Length 284;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YRRFVSVR 10
Db 81 YKRFDVGVK 90
|||||:|

RESULT 7
O80624
ID O80624 PRELIMINARY; PRT; 328 AA.

```

```

AC O80624;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE F12L6.3 PROTEIN.
GN F12L6.3.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC004218; AAC27828.1; -
SQ SEQUENCE 328 AA; 37138 MW; E258DA27FA7F0F3C CRC64;

Query Match 71.4%; Score 35; DB 10; Length 328;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRRFVSVR 10
Db 232 HRRFSVSMR 241
|||||:|

RESULT 8
Q9PPS7
ID Q9PPS7 PRELIMINARY; PRT; 146 AA.
AC Q9PPS7
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN UU563.
GN UU563.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002154; AAF30976.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 146 AA; 17597 MW; 01678E52C8CBF55E CRC64;

Query Match 69.4%; Score 34; DB 2; Length 146;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YRRFVSVR 10
Db 70 YRREAVKLR 79
|||||:|

RESULT 9
Q9KKM9
ID Q9KKM9 PRELIMINARY; PRT; 287 AA.
AC Q9KKM9
DT 01-OCT-2000 (Tremblrel. 15, Created)

```

Wed Feb 13 07:52:05 2002

```

01-OCT-2000 (TREMBlrel. 15, Last sequence update)
01-JUN-2001 (TREMBlrel. 17, Last annotation update)
TRANSCRIPTIONAL REGULATOR, ARAC/XYLS FAMILY.
VCA1074.
GN Vibrio cholerae.
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC NCBI_TaxID=686;
OX [1]
RN
SEQUENCE FROM N.A.
RP STRAIN=EL TOR N16961 / SEROTYPE O1;
RC MEDLINE=20406833; PubMed=10952301;
RX Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholatae."
RL Nature 406:477-483(2000).
CC -1- SIMILARITY: BELONGS TO THE ARAC/XYLS FAMILY OF TRANSCRIPTIONAL
REGULATORS.
CC EMBL; AE004433; AAF96967.1; -.
DR TIGR; VCA1074; -.
DR InterPro; IPR000005; HTHARAC.
DR Pfam; PF00165; HTH_ARAC; 1.
DR PRINTS; PR00032; HTHARAC.
DR SMART; SM00342; HTH_ARAC; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
DR Complete proteome; DNA-binding; Transcription regulation.
KW Complete proteome; 31460 MW; 7FD5FB18E8C08B5 CRC64;
SQ SEQUENCE 287 AA; 41460 MW; 7FD5FB18E8C08B5 CRC64;

Query Match 69.4%; Score 34; DB 2; Length 287;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YRRRFSVS 9
   1:|||||:
DB 279 YKRRFSSSI 287

RESULT 10
ID Q9CJX3 PRELIMINARY; PRT; 437 AA.
AC Q9CJX3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN PM1866.
GN PM1866.
OS Pasteurella multocida.
OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Bacteria.
OC Pasteurella.
OX NCBI_TaxID=747;
OX [1]
RN
SEQUENCE FROM N.A.
RP STRAIN=PM70.
RC MEDLINE=21145866; PubMed=11248100;
RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida Pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006224; AAK03950.1; -.
DR InterPro; IPR002792; DUF90.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR001566; TRMA_1.
DR Pfam; PF01938; DUF90; 1.
DR PROSITE; PS01230; TRMA_1; UNKNOWN_1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 437 AA; 50021 MW; 03D4876D3533EBC1 CRC64;

Query Match 69.4%; Score 34; DB 2; Length 437;
Best Local Similarity 70.0%; Pred. No. 77;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YRRRFSVS 10
   1:||||:
DB 132 YRRRRLSVR 141

RESULT 11
ID Q9X6Y5 PRELIMINARY; PRT; 1023 AA.
AC Q9X6Y5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE BETA-GALACTOSIDASE.
GN LACZ.
OS Bifidobacterium longum.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=1679;
OX [1]
RN
SEQUENCE FROM N.A.
RP STRAIN=MB219;
RC Rossi M., Altomare L., Brigidi P., Gonzalez-Vara A., Matteuzzi D.;
RA "Nucleotide sequence, expression and transcriptional analysis of the
RT Bifidobacterium longum MB219 lacZ gene."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ242596; CAB44428.1; -.
DR HSSP; P00722; 1BGL.
DR InterPro; IPR001649; Glyco_hydro_2.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR SEQUENCE 1023 AA; 114497 MW; 9B8BEB4786441B60 CRC64;

Query Match 69.4%; Score 34; DB 2; Length 1023;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YRRRFSVS 8
   1:||||:
DB 128 YRRKFTVS 135

RESULT 12
ID Q9L6Q1 PRELIMINARY; PRT; 135 AA.
AC Q9L6Q1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE STM1.49 PROTEIN.
GN STM1.49.
OS Salmonella typhimurium LT2.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteria.
OC Salmonella.
OX NCBI_TaxID=99287;
OX [1]
RN
SEQUENCE FROM N.A.
RP STRAIN=SGSC1412;
RC Washu;
RA "The Salmonella typhimurium Genome Sequencing Project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=SGSC1412;
RC Waterston R.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF233324; AAF33451.1; -.
SQ SEQUENCE 437 AA; 50021 MW; 03D4876D3533EBC1 CRC64;

```

SQ SEQUENCE 125 AA; 14296 MW; F2871B309FAA0180 CRC64;

Query Match 67.3%; Score 33; DB 2; Length 125;  
Best Local Similarity 75.0%; Pred. No. 34;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YRRRFSVS 8  
      |||||  
DB 3 YRRRFEVA 10

## RESULT 13

ID O29512 PRELIMINARY; PRT; 249 AA.  
AC O29512;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
DE CONSERVED HYPOTHETICAL PROTEIN.  
GN AF0746.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
OC Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Richardson K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richmond D.L., Kierlavage A.R., Graham D.E., Kyriplides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
DR EMBL; AE001053; AAB90493.1; .  
DR TIGR; AF0746; .  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 249 AA; 27620 MW; 36D4DCFF0501C9B CRC64;

Query Match 67.3%; Score 33; DB 1; Length 249;  
Best Local Similarity 66.7%; Pred. No. 69;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 YRRRFSVS 10  
      |||||  
DB 71 RRRFAVNR 79

## RESULT 14

O9XGP7 PRELIMINARY; PRT; 252 AA.  
ID O9XGP7;  
AC O9XGP7;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE ESTS AU030740(E60171).  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC  
RT clone: P0026F07.";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF000364; BAA1774.1; .  
DR InterPro; IPR000051; SAM\_bind.  
DR InterPro; IPR002935; Methyltransf\_3.  
DR Pfam; PF01596; Methyltransf\_3; 1.  
SQ SEQUENCE 252 AA; 27771 MW; E4FC076B47483926 CRC64;

Query Match 67.3%; Score 33; DB 10; Length 252;  
Best Local Similarity 75.0%; Pred. No. 70;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 RRRFSVSR 10  
      |||||  
DB 215 RRRFSVAIR 222

## RESULT 15

O9IBW2 PRELIMINARY; PRT; 513 AA.  
ID O9IBW2;  
AC O9IBW2;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE PROTEIN KINASE.  
GN UL13.  
OS Turkey herpesvirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.  
OX NCBI\_TaxID=10390;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GA;  
RX MEDLINE=92237304; PubMed=1315048;  
RA Jones D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;  
RT "Marek disease virus encodes a basic-leucine zipper gene resembling  
RT the foV/jun oncogenes that is highly expressed in lymphoblastoid  
tumors.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GA;  
RA Lee L.F., Wu P., Sui D., Ren D., Kung H.J.;  
RT "The Complete OL Sequence of Serotype I Marek's Disease Virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).  
DR EMBL; AF147806; AAF66749.1; .  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; UNKNOWN1.  
KW ATP-binding; Kinase; Transferase.  
SQ SEQUENCE 513 AA; 58905 MW; D73D04608C32DC07 CRC64;

Query Match 67.3%; Score 33; DB 12; Length 513;  
Best Local Similarity 60.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 YRRRFSVSR 10  
      |||||  
DB 77 HRRRFSNIR 86

Search completed: February 12, 2002, 12:38:39  
Job time: 752 sec

us-09-485-571-22.rspt

Wed Feb 13 07:52:05 2002

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:32 ; Search time 242.57 seconds  
(without alignments)  
3.054 Million cell updates/sec

Title: US-09-485-571-23  
Perfect score: 50  
Sequence: 1 RRLSYRRRF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	10	20	AAW99410
2	50	100.0	10	21	AAW93189
3	50	100.0	10	21	AAW93618
4	50	100.0	15	20	AAW99411
5	45	90.0	18	20	AAW99412
6	45	90.0	18	20	AAW99403
7	45	90.0	18	21	AAW93177
8	45	90.0	18	21	AAW93179
9	45	90.0	18	21	AAW93615
10	45	90.0	18	21	AAW93616
11	40	80.0	14	18	AAW36264

12	40	80.0	16	18	AAW36279	Antimicrobial prot
13	40	80.0	17	18	AAW36440	Antimicrobial prot
14	39	78.0	18	16	AAW78776	Protegrin peptide
15	39	78.0	18	18	AAW18153	Cationic, antimicr
16	38	76.0	304	22	AAW92371	C glutamicum proce
17	37	74.0	16	16	AAW78768	Protegrin peptide
18	37	74.0	18	18	AAW36285	Antimicrobial prot
19	37	74.0	18	18	AAW36429	Antimicrobial prot
20	37	74.0	18	18	AAW18150	Cationic, antimicr
21	37	74.0	18	18	AAW09084	Cationic, antimicr
22	37	74.0	18	18	AAW09085	Cationic, antimicr
23	36	72.0	16	16	AAW78766	Protegrin peptide
24	36	72.0	1045	20	AAW07482	Human chondrocyte-
25	36	72.0	1045	21	AAW91947	Human cytoskeleton
26	35	70.0	13	18	AAW36265	Antimicrobial prot
27	35	70.0	13	18	AAW36438	Antimicrobial prot
28	35	70.0	13	18	AAW35602	Antimicrobial pept
29	35	70.0	14	18	AAW36272	Antimicrobial prot
30	35	70.0	14	18	AAW36263	Antimicrobial prot
31	35	70.0	15	18	AAW36280	Antimicrobial prot
32	35	70.0	15	18	AAW36261	Antimicrobial prot
33	35	70.0	15	18	AAW36262	Antimicrobial prot
34	35	70.0	15	18	AAW36268	Antimicrobial prot
35	35	70.0	15	18	AAW35599	Antimicrobial pept
36	35	70.0	16	16	AAW78752	Protegrin PG-2. S
37	35	70.0	16	16	AAW78755	Protegrin peptide
38	35	70.0	16	16	AAW78756	Protegrin peptide
39	35	70.0	16	18	AAW36253	Antimicrobial prot
40	35	70.0	16	18	AAW36278	Antimicrobial prot
41	35	70.0	16	18	AAW36282	Antimicrobial prot
42	35	70.0	16	18	AAW36270	Antimicrobial prot
43	35	70.0	16	18	AAW18124	Cationic, antimicr
44	35	70.0	16	18	AAW09080	Cationic, antimicr
45	35	70.0	16	18	AAW18118	Cationic, antimicr

#### ALIGNMENTS

RESULT 1  
AAW99410  
ID AAW99410 standard; peptide; 10 AA.  
XX  
AC AAW99410;  
XX  
DT 08-JUN-1999 (first entry)  
XX  
DE Protegrin derivative peptide SM2195.  
XX  
KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.  
XX  
OS Synthetic.  
XX  
PN WO9907728-A2.  
XX  
PD 18-FEB-1999.  
XX  
PF 05-AUG-1998; 98WO-FR01757.  
XX  
PR 12-AUG-1997; 97FR-0010297.  
XX  
PA (SYNT-) SYNT:EM SA.  
XX  
PI Calas B, Chavanieu A, Grassy G, Kaczorek M;  
XX  
DR WPI; 1999-190034/16.  
XX  
PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
PT as carriers to deliver active agents into cells  
XX

are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must be Trp; each B is aa containing a side chain that includes a basic group; and each X is an aliphatic or aromatic aa. The linear peptide may be retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment containing at least 5, preferably at least 7 consecutive aa from (a)-(c). Peptides able to cross the BBB include protegrins, Antennapedia, tachypleins, transportin, etc. Of these several families have cytolytic effects and are termed peptide antibiotics. They fall into 3 main categories based on their structure: (i) peptides with alpha-helices, e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked beta-sheets, e.g. protegrin, tachypleins, defensins; (iii) peptides with no major structure but containing bends due to the presence of pro residues, e.g. bactericins and PR39. The peptides of the invention fall into the peptide antibiotic categories defined above: (a)-peptides are based on the Antennapedia family peptides; (b)-peptides are based on protegrins; and (c)-peptides are based on tachypleins. This sequence represents a synthetic linear peptide designed on peptides able to cross the BBB and is conjugated to a doxorubicin molecule by a succinate linker. Conjugates of the linear peptides and the active agent are particularly used to treat, prevent or diagnose brain cancer, Alzheimer's or Parkinson's diseases, depression, pain and meningitis, but also for studying drug behaviour in BBB models.

Sequence 10 AA;

Query Match 100.0%; Score 50; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.002;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLSYRRRRF 10  
Db 1 RRLSYRRRRF 10

RESULT 3  
AAAY93618  
ID AAY93618 standard; peptide; 10 AA.

XX AC AAY93618;  
XX DT 25-SEP-2000 (first entry)

XX DE Peptide which may be linked to anticancer agents.

XX KW Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
XX KW cancer.

XX OS Unidentified.

XX PN WO200032237-A1.

XX PD 08-JUN-2000.

XX PF 26-NOV-1999; 99WO-FR02939.

XX PR 30-NOV-1998; 98FR-0015073.

XX PA (SYNT-) SYNT:EM SA.

XX PI Tamsamani J, Kaczorek M, Colin De Verdiere A;

XX DR WPI; 2000-412166/35.

XX PT New composition useful for cancer treatment and prevention, contains anticancer agent and peptide vector that transports agent into cells

XX PS Disclosure; Page 8; 34pp; French.

XX CC The specification describes a pharmaceutical composition, which comprises at least one anticancer agent associated with at least one peptide that can transport it into cancer cells and which inhibits

PS Claim 7; Page 28; 37pp; French.  
XX This peptide represents a linear derivative of the protegrin family of peptide antibiotics. Protegrin antibiotics form part of the peptide antibiotic family which contain a beta-sheet secondary structure linked by disulphide bridges. The new derivatives are linear and lack the disulphide bridge. The novel derivatives are used to deliver active agents to an organism, e.g. therapeutic proteins, antibodies (or their fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents, antivirals and anti-inflammatories, etc. The derivatives are non-toxic and non-lytic but can cross mammalian cell membranes rapidly by a passive mechanism, so can deliver active agents to cytoplasm and nucleus, including crossing the blood-brain barrier.

XX Sequence 10 AA;

Query Match 100.0%; Score 50; DB 20; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.002;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLSYRRRRF 10  
Db 1 RRLSYRRRRF 10

RESULT 2  
AAAY93189  
ID AAY93189 standard; peptide; 10 AA.

XX AC AAY93189;

XX DT 06-DEC-2000 (first entry)

XX DE Protegrin-like peptide antibiotic Doxo-SynB3.

XX KW Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
XX KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
XX KW Antennapedia; tachyplein; peptide antibiotic; Alzheimer's disease;  
XX KW cancer; Parkinson's disease; depression; pain; meningitis.

XX OS Synthetic.

XX PH Key Location/Qualifiers

XX FT Modified-site 1 /note= "linked to doxorubicin via a succinate  
XX FT (-CO-(CH2)2-CO- linker"

XX PN WO200032236-A1.

XX PD 08-JUN-2000.

XX PF 26-NOV-1999; 99WO-FR02938.

XX PR 30-NOV-1998; 98FR-0015074.

XX PA (SYNT-) SYNT:EM SA.

XX PI Clair P, Kaczorek M, Tamsamani J;

XX DR WPI; 2000-422871/36.

XX PT Use of linear peptides as vectors for active ingredients, useful for diagnosis and treatment of central nervous system diseases, can transport agents passively across the blood-brain barrier

XX PS Example III; Page 22; 54pp; French.

XX CC The invention relates to the use of linear peptides, coupled to an active agent, to prepare a composition able to cross the blood-brain barrier for diagnosis or treatment of disorders localised in the central nervous system. The linear peptide preferably has the formula: (a) X1-X16; (b); BXXBXXXXBXXXXB; or (c) BXXBXXXXBXXXXB, where: each of X1-X16

CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.

XX Sequence 10 AA;

Query Match 100.0%; Score 50; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.002;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLSYSRRRF 10  
 Db 1 rrlsyrrrf 10

RESULT 4

AAW99411  
 ID AAW99411 standard; peptide; 15 AA.

XX AC AAW99411;

XX DT 08-JUN-1999 (first entry)

XX DE Protegrin derivative peptide SM2193.

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX OS Synthetic.

XX PN WO9907728-A2.

XX PD 18-FEB-1999.

XX PF 06-AUG-1998; 98WO-FR01757.

XX PR 12-AUG-1997; 97FR-0010297.

XX PA (SYNT-) SYNT:EM SA.

XX PI Calas B, Chavanieu A, Grassy G, Kaczorek M;

XX DR WPI; 1999-190034/16.

XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX Claim 7; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 15 AA;

Query Match 100.0%; Score 50; DB 20; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.003;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLSYSRRRF 10  
 Db 1 rrlsyrrrf 10

RESULT 5

AAW99412  
 ID AAW99412 standard; peptide; 18 AA.

XX AC AAW99412;

XX DT 08-JUN-1999 (first entry)

XX DE Protegrin derivative peptide SM2196.

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX OS Synthetic.

XX PN WO9907728-A2.

XX PD 18-FEB-1999.

XX PF 06-AUG-1998; 98WO-FR01757.

XX PR 12-AUG-1997; 97FR-0010297.

XX PA (SYNT-) SYNT:EM SA.

XX PI Calas B, Chavanieu A, Grassy G, Kaczorek M;

XX DR WPI; 1999-190034/16.

XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX Claim 7; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 18 AA;

Query Match 90.0%; Score 45; DB 20; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 0.032;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
 Db 4 rlsysrrrf 12

RESULT 6

AAW99403

ID AAW99403 standard; peptide; 18 AA.

XX

Wed Feb 13 07:52:06 2002

AAW99403;  
08-JUN-1999 (first entry)  
Protegrin derivative peptide SM1738.  
Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
nucleus; blood-brain barrier.  
Synthetic.  
WO9907728-A2.  
18-FEB-1999.  
06-AUG-1998; 98WO-FR01757.  
12-AUG-1997; 97FR-0010297.  
(SYNT-) SYNT:EM SA.  
Calas B, Chavanieu A, Grassy G, Kaczorek M;  
WPI; 1999-190034/16.  
Derivatives of antibiotic peptides lacking disulfide bridges - used  
as carriers to deliver active agents into cells  
Claim 7; Page 28; 37pp; French.  
This peptide represents a linear derivative of the protegrin family of  
peptide antibiotics. Protegrin antibiotics form part of the peptide  
antibiotic family which contain a beta-sheet secondary structure linked  
by disulphide bridges. The new derivatives are linear and lack the  
disulphide bridge. The novel derivatives are used to deliver active  
agents to an organism, e.g. therapeutic proteins, antibodies (or their  
fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
antivirals, and anti-inflammatory, etc. The derivatives are non-toxic  
and non-lytic but can cross mammalian cell membranes rapidly by a passive  
mechanism, so can deliver active agents to cytoplasm and nucleus,  
including crossing the blood-brain barrier.  
Sequence 18 AA;  
Query Match 90.0%; Score 45; DB 20; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.032;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 RLSYSRRRF 10  
Db 4 rlsysrrrf 12  
RESULT 7  
AAW93177  
ID AAY93177 standard; peptide; 18 AA.  
AC AAY93177;  
XX  
XX 06-DEC-2000 (first entry)  
XX  
XX Protegrin-like peptide antibiotic Doxo-SynBI.  
XX  
XX Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
XX blood-brain barrier; diagnostic; central nervous system; protegrin;  
XX Antennapedia; tachyplein; peptide antibiotic; Alzheimer's disease;  
XX cancer; Parkinson's disease; depression; pain; meningitis.  
XX  
XX Synthetic.  
XX

FH Key Location/Qualifiers  
FT Modified-site 1  
FT /note= "linked to doxorubicin via a succinate  
XX (-CO-(CH2)2-CO-) linker; optionally linked  
XX to benzylpenicillin by a glycoamide linker".  
XX  
XX WO200032236-A1.  
XX 08-JUN-2000.  
XX 26-NOV-1999; 99WO-FR02938.  
XX 30-NOV-1998; 98FR-0015074.  
XX (SYNT-) SYNT:EM SA.  
XX Clair P, Kaczorek M, Tamsamani J;  
XX WPI; 2000-422871/36.  
XX  
XX Use of linear peptides as vectors for active ingredients, useful for  
XX diagnosis and treatment of central nervous system diseases, can  
XX transport agents passively across the blood-brain barrier  
XX  
XX Example 1; Page 13; 54pp; French.  
XX  
XX The invention relates to the use of linear peptides, coupled to an active  
XX agent, to prepare a composition able to cross the blood-brain barrier  
XX for diagnosis or treatment of disorders localised in the central nervous  
XX system. The linear peptide preferably has the formula: (a) X1-X16  
XX (b); BXXXXXXXBXXXXXB; or (c) BXXXXXXXBXXXXXB, where: each of X1-X16  
XX are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
XX be Trp; each B is aa containing a side chain that includes a basic group;  
XX and each X is an aliphatic or aromatic aa. The linear peptide may be  
XX containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
XX peptides able to cross the BBB include protegrins, Antennapedia,  
XX tachypleins, transportin, etc. Of these several families have cytolytic  
XX effects and are termed peptide antibiotics. They fall into 3 main  
XX categories based on their structure: (i) peptides with alpha-helices,  
XX e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
XX beta-sheets, e.g. protegrin, tachypleins, defensins; (iii) peptides  
XX with no major structure but containing bends due to the presence of  
XX pro residues, e.g. bactericins and PR39. The peptides of the invention  
XX fall into the peptide antibiotic categories defined above: (a)-peptides  
XX are based on the Antennapedia family peptides; (b)-peptides are based on  
XX protegrins; and (c)-peptides are based on tachypleins. This sequence  
XX represents a synthetic linear peptide designed on peptides able to cross  
XX the BBB and is conjugated to a doxorubicin molecule by a succinate  
XX linker. The peptide may also be linked to a benzylpenicillin molecule  
XX by a glycoamide linker.  
XX Conjugates of the linear peptides and the active agent are particularly  
XX used to treat, prevent or diagnose brain cancer, Alzheimer's or  
XX Parkinson's diseases, depression, pain and meningitis, but also for  
XX studying drug behaviour in BBB models.  
XX  
XX Sequence 18 AA;  
XX  
XX Query Match 90.0%; Score 45; DB 21; Length 18;  
XX Best Local Similarity 100.0%; Pred. No. 0.032;  
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 2 RLSYSRRRF 10  
XX Db 4 rlsysrrrf 12  
XX  
XX RESULT 8  
XX AAY93179  
XX ID AAY93179 standard; peptide; 18 AA.  
XX  
XX AC AAY93179;  
XX

XX 06-DEC-2000 (first entry)  
DE Protegrin-like peptide antibiotic Dal-SynBI.  
XX  
XX Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
KW cancer; Parkinson's disease; depression; pain; meningitis; dalargin.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
FT Cross-links I  
FT /note= "cross-links to a molecule of dalargin via  
FT a disulphide linker"  
XX  
XX WO200032236-A1.  
XX  
XX 08-JUN-2000.  
XX  
XX 26-NOV-1999; 99WO-FR02938.  
XX  
XX 30-NOV-1998; 98FR-0015074.  
XX  
XX (SYNT-) SYNT:EM SA.  
XX  
XX Clair P, Kaczorek M, Tamsamani J;  
PI WPI; 2000-422871/36.  
XX  
XX Use of linear peptides as vectors for active ingredients, useful for  
PT diagnosis and treatment of central nervous system diseases, can  
PT transport agents passively across the blood-brain barrier.  
XX  
XX Example II; Page 20; 54pp; French.  
XX  
XX The invention relates to the use of linear peptides, coupled to an active  
CC agent, to prepare a composition able to cross the blood-brain barrier  
CC for diagnosis or treatment of disorders localised in the central nervous  
CC system. The linear peptide preferably has the formula: (a) X1- X16;  
CC (b); BXXBXXXXBXXXXXB; or (c) BXXBXXXXBXXXXXBXB, where: each of X1-X16  
CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
CC be Trp; each B is aa containing a side chain that includes a basic group;  
CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
CC Peptides able to cross the BBB include protegrins, Antennapedia,  
CC tachyplesins, transportan, etc. Of these several families have cytolytic  
CC effects and are termed peptide antibiotics. They fall into 3 main  
CC categories based on their structure: (i) peptides with alpha-helices,  
CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
CC beta-sheets, e.g. protegrin, tachyplesins, defensins; (iii) peptides  
CC with no major structure but containing bends due to the presence of  
CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
CC fall into the peptide antibiotic categories defined above: (a)-peptides  
CC are based on the Antennapedia family peptides; (b)-peptides are based on  
CC protegrins; and (c)-peptides are based on tachyplesins. This sequence  
CC represents a synthetic linear peptide designed on peptides able to cross  
CC the BBB and is conjugated to a dalargin molecule by a disulphide linker.  
CC Conjugates of the linear peptides and the active agent are particularly  
CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
CC Parkinson's diseases, depression, pain and meningitis, but also for  
CC studying drug behaviour in BBB models.  
XX  
XX Sequence 18 AA:  
SQ  
Query Match 90.0%; Score 45; DB 21; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.032;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 RLSYSRRRF 10  
DE Peptide which may be linked to anticancer agents.

Db 4 rlsysrrrf 12  
RESULT 9  
AA93615  
ID AA93615 standard; peptide; 18 AA.  
XX AC AA93615;  
XX 25-SEP-2000 (first entry)  
DT Peptide which may be linked to anticancer agents.  
DE Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
XX cancer.  
KW Unidentified.  
XX WO200032237-A1.  
XX 08-JUN-2000.  
XX 26-NOV-1999; 99WO-FR02939.  
XX 30-NOV-1998; 98FR-0015073.  
XX (SYNT-) SYNT:EM SA.  
XX Tamsamani J, Kaczorek M, Colin De Verdlere A;  
PI WPI; 2000-412166/35.  
XX New composition useful for cancer treatment and prevention, contains  
PT anticancer agent and peptide vector that transports agent into cells  
XX  
XX Disclosure; Page 8; 34pp; French.  
XX  
XX The specification describes a pharmaceutical composition, which  
CC comprises at least one anticancer agent associated with at least one  
CC peptide that can transport it into cancer cells and which inhibits  
CC development of resistance to the anticancer agent. By using the  
CC peptide as a vector for delivery of the anticancer agent, mechanisms  
CC that cause cancer cells to become resistant to the agent, particularly  
CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
CC produced by chemical synthesis, can be coupled easily to the agent,  
CC cross mammalian cell membranes rapidly by a passive mechanism (no  
CC receptors required), and are non-toxic and non-lytic. The compositions  
CC are used to treat cancer. The present sequence represents a peptide  
CC which may be linked to the anticancer agents of the invention.  
XX  
XX Sequence 18 AA:  
SQ  
Query Match 90.0%; Score 45; DB 21; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.032;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 RLSYSRRRF 10  
DE Peptide which may be linked to anticancer agents.

Wed Feb 13 07:52:06 2002

us-09-485-571-23.rag

KW Anticancer agent; cancer cell; resistance: P-glycoprotein pump;  
 KW cancer.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200032237-A1.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 26-NOV-1999; 99WO-FR02939.  
 XX  
 PR 30-NOV-1998; 98FR-0015073.  
 XX  
 PA (SYNT-) SYNT:EM SA.  
 XX  
 PI Tamsamani J, Kaczorek M, Colin De Verdiere A;  
 XX  
 DR WPI; 2000-412166/35.  
 XX  
 PT New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells  
 XX  
 PS Disclosure; Page 8; 34pp; French.  
 XX  
 CC The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can translocate it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.  
 XX  
 SQ Sequence 18 AA;  
 XX  
 Query Match 90.0%; Score 45; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 0.032;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 RLSSRRRF 10  
 Db 4 rlsysrrrf 12  
 RESULT 11  
 ID AAW36264 standard; peptide; 14 AA.  
 AC AAW36264;  
 XX  
 DT 13-FEB-1998 (first entry)  
 DE Antimicrobial proteogrin peptide PC34 (64).  
 XX  
 KW Antimicrobial proteogrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
 XX  
 OS Synthetic.  
 OS -fus scrofa.

XX WO9718826-A1.  
 XX  
 PD 29-MAY-1997.  
 XX  
 PF 22-NOV-1996; 96WO-US18544.  
 XX  
 PR 21-NOV-1996; 96US-0752852.  
 PR 22-NOV-1995; 95US-0562346.  
 PR 17-MAY-1996; 96US-0649811.  
 PR 01-AUG-1996; 96US-0690921.  
 XX  
 PA (INTR-) INTRABIOTICS PHARM INC.  
 PA (REGC) UNIV CALIFORNIA.  
 XX  
 PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
 XX  
 DR WPI; 1997-297871/27.  
 XX  
 PT New antimicrobial proteogrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)  
 XX  
 PS Claim 23; Page 106; 130pp; English.  
 XX  
 CC The present sequence is an antimicrobial proteogrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections.  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX  
 SQ Sequence 14 AA;  
 XX  
 Query Match 80.0%; Score 40; DB 18; Length 14;  
 Best Local Similarity 80.0%; Pred. No. 0.22;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 RRLSSRRRF 10  
 Db 1 rrlsycrrrf 10  
 RESULT 12  
 ID AAW36279 standard; peptide; 16 AA.  
 XX  
 AC AAW36279;  
 XX  
 DT 13-FEB-1998 (first entry)  
 DE Antimicrobial proteogrin peptide PC34a (79).  
 XX  
 KW Antimicrobial proteogrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW

KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
 XX  
 OS Synthetic.  
 OS Sus scrofa.  
 XX  
 PN W09718826-A1.  
 XX  
 PD 29-MAY-1997.  
 XX  
 PF 22-NOV-1996; 96WO-US18544.  
 XX  
 PR 21-NOV-1996; 96US-0752852.  
 PR 21-NOV-1996; 96US-0562346.  
 PR 17-MAY-1996; 96US-0649811.  
 PR 01-AUG-1996; 96US-0690921.  
 XX  
 PA (INTR-) INTRABIOTICS PHARM INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
 XX  
 DR WPI; 1997-297871/27.  
 XX  
 PT New antimicrobial protegrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)  
 XX  
 PS Claim 23; Page 106; 130pp; English.  
 XX  
 CC The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets, which  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX  
 SQ Sequence 16 AA;

Query Match 80.0%; Score 40; DB 18; Length 16;  
 Best Local Similarity 80.0%; Pred. No. 0.26;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRLSYRRRRF 10  
 ||| | |||  
 DB 1 rrlcyrrrf 10

RESULT 13  
 ID AAW36440  
 AC AAW36440 standard; peptide: 17 AA.  
 XX  
 AC AAW36440;  
 XX  
 DT 13-FEB-1998 (first entry)  
 XX

DE Antimicrobial protegrin peptide (240).  
 XX  
 KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
 XX  
 OS Synthetic.  
 OS Sus scrofa.  
 XX  
 PN W09713826-A1.  
 XX  
 PD 29-MAY-1997.  
 XX  
 PF 22-NOV-1996; 96WO-US18544.  
 XX  
 PR 21-NOV-1996; 96US-0752852.  
 PR 21-NOV-1996; 96US-0562346.  
 PR 17-MAY-1996; 96US-0649811.  
 PR 01-AUG-1996; 96US-0690921.  
 XX  
 PA (INTR-) INTRABIOTICS PHARM INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
 XX  
 DR WPI; 1997-297871/27.  
 XX  
 PT New antimicrobial protegrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)  
 XX  
 PS Claim 23; Page 111; 130pp; English.  
 XX  
 CC The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX  
 SQ Sequence 17 AA;

Query Match 80.0%; Score 40; DB 18; Length 17;  
 Best Local Similarity 80.0%; Pred. No. 0.27;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

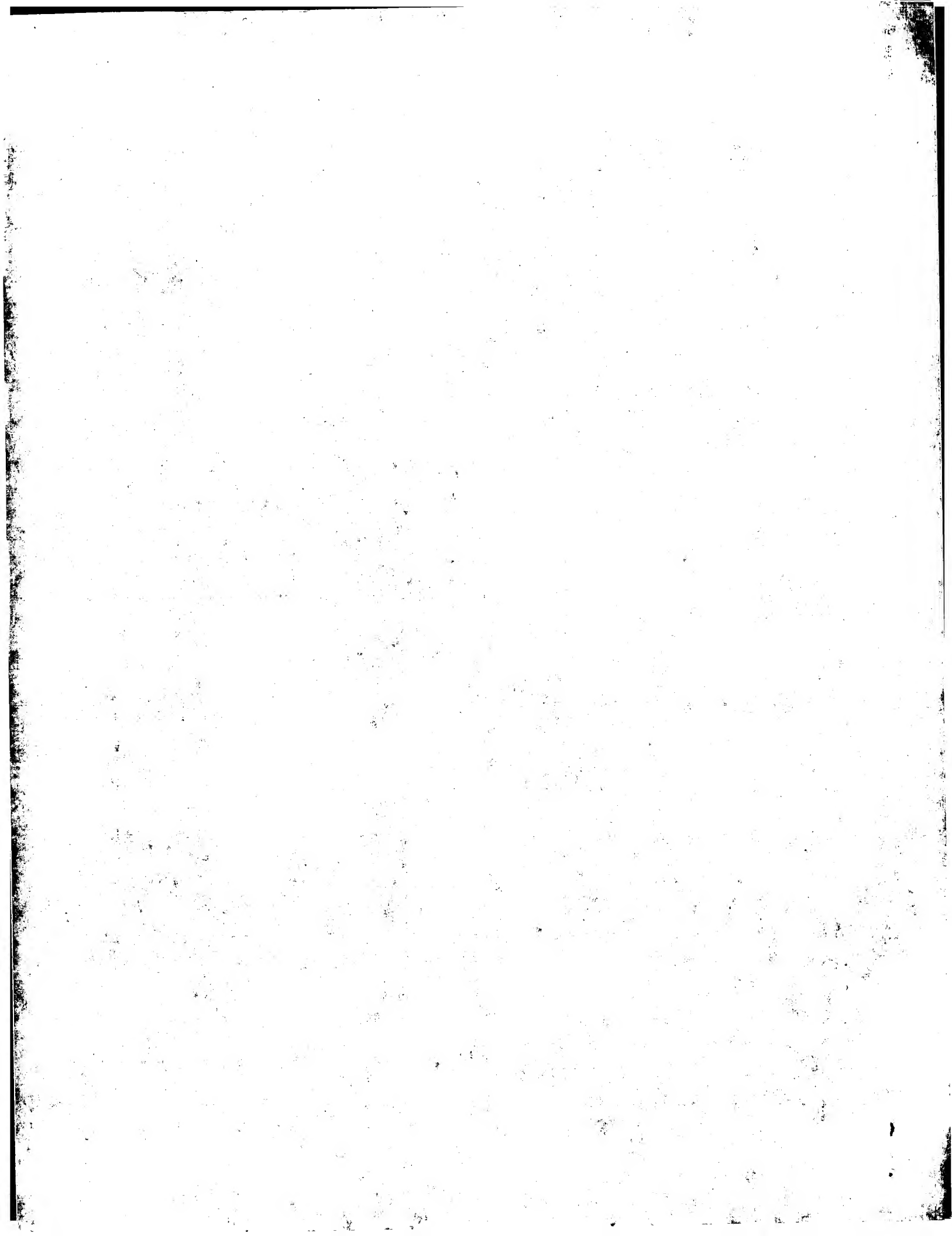
OY 1 RRLSYRRRRF 10  
 ||| | |||  
 DB 2 rrlcyrrrf 11

XX  
DT 11-AUG-1997 (first entry)

search completed: February 12, 2002, 12:30:32



Job time: 365 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:32 ; Search time 242.57 seconds  
(without alignments)  
4.581 Million cell updates/sec

Title: US-09-485-571-24

Perfect score: 71

Sequence: 1 RRLSYRRRFSVSVR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	71	100.0	15	20	AAW99411	Protegrin derivati
2	61	85.9	18	20	AAW99403	Protegrin derivati
3	61	85.9	18	21	AAW93616	Peptide which may
4	55	77.5	18	16	AAW78776	Protegrin peptide
5	53	74.6	18	20	AAW99412	Protegrin derivati
6	53	74.6	18	21	AAW93177	Protegrin-like pep
7	53	74.6	18	21	AAW93179	Protegrin-like pep
8	53	74.6	18	21	AAW93615	Peptide which may
9	50	70.4	10	20	AAW99410	Protegrin derivati
10	50	70.4	10	21	AAW93189	Protegrin-like pep
11	50	70.4	10	21	AAW93618	Peptide which may

12	46	64.8	14	18	AAW36264	Antimicrobial prot
13	46	64.8	16	18	AAW36279	Antimicrobial prot
14	46	64.8	17	18	AAW36440	Antimicrobial prot
15	46	64.8	17	18	AAW09078	Cationic, antimicr
16	45	63.4	18	18	AAW36285	Antimicrobial prot
17	45	63.4	18	18	AAW36429	Antimicrobial prot
18	45	63.4	18	18	AAW18153	Cationic, antimicr
19	45	63.4	18	18	AAW09084	Cationic, antimicr
20	45	63.4	18	18	AAW09085	Cationic, antimicr
21	44	62.0	16	16	AAW78768	Protegrin peptide
22	44	62.0	18	18	AAW18151	Cationic, antimicr
23	44	62.0	18	18	AAW18152	Cationic, antimicr
24	44	62.0	526	21	AAW12717	Streptococcus pneu
25	43	60.6	14	18	AAW36318	Antimicrobial prot
26	43	60.6	1045	20	AAW07482	Human chondrocyte-
27	43	60.6	1045	21	AAW91947	Human cytoskeleton
28	42.5	59.9	304	22	AAW92371	C glutamylum prote
29	42	59.2	10	20	AAW99409	Protegrin derivati
30	42	59.2	198	21	AAW42941	Human ORFX ORF2705
31	41	57.7	13	18	AAW36265	Antimicrobial prot
32	41	57.7	13	18	AAW36438	Antimicrobial prot
33	41	57.7	13	18	AAW35602	Antimicrobial pept
34	41	57.7	14	18	AAW36220	Antimicrobial prot
35	41	57.7	14	18	AAW36263	Antimicrobial prot
36	41	57.7	15	18	AAW36280	Antimicrobial prot
37	41	57.7	15	18	AAW36261	Antimicrobial prot
38	41	57.7	15	18	AAW36262	Antimicrobial prot
39	41	57.7	16	16	AAW78755	Protegrin peptide
40	41	57.7	16	16	AAW78756	Protegrin peptide
41	41	57.7	16	16	AAW78766	Protegrin peptide
42	41	57.7	16	18	AAW36278	Antimicrobial prot
43	41	57.7	16	18	AAW36270	Antimicrobial prot
44	41	57.7	17	18	AAW36276	Antimicrobial prot
45	41	57.7	17	18	AAW36277	Antimicrobial prot

#### ALIGNMENTS

RESULT 1  
AAW99411 standard; peptide; 15 AA.  
XX  
AC AAW99411;  
XX  
DT 08-JUN-1999 (first entry)  
XX  
DE Protegrin derivative peptide SM2193.  
XX  
KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.  
XX  
OS Synthetic.  
XX  
PN WO9907728-A2.  
XX  
PD 18-FEB-1999.  
XX  
PF 06-AUG-1998; 98WO-FR01757.  
XX  
PR 12-AUG-1997; 97FR-0010297.  
XX  
(SYNT-) SYNT:EM SA.  
XX  
PI Calas B. Chavanieu A, Grassy G, Kaczorek M;  
XX  
DR WPI; 1999-190034/16.  
XX  
PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
XX as carriers to deliver active agents into cells

This peptide represents a linear derivative of the protegrin family of peptide antibiotics. Protegrin antibiotics form part of the peptide antibiotic family which contain a beta-sheet secondary structure linked by disulphide bridges. The new derivatives are linear and lack the disulphide bridge. The novel derivatives are used to deliver active agents to an organism, e.g. therapeutic proteins, antibodies (or their fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents, antivirals and anti-inflammatories, etc. The derivatives are non-toxic and non-lytic but can cross mammalian cell membranes rapidly by a passive mechanism, so can deliver active agents to cytoplasm and nucleus, including crossing the blood-brain barrier.

Query Match 85.9%; Score 61; DB 21; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

	QY	14
2	RLSYRRRFSVSV	14
		11
4	rlsvrrrrfsvsv	16

## RESULT

"linked to doxorubicin via a succinate (-CO-(CH<sub>2</sub>)<sub>2</sub>-CO-) linker; optionally linked

to benzylpenicillin by a glycoamide linker"

FT XX WO200032236-A1.  
 PN XX 08-JUN-2000.  
 PD XX  
 XX XX  
 XX XX 26-NOV-1999; 99WO-FR02938.  
 PF XX  
 XX XX 30-NOV-1998; 98FR-0015074.  
 PR XX  
 XX XX (SYNT-) SYNT:EM SA.  
 PA XX  
 XX XX  
 PI XX Clair P, Kaczorek M, Tamsamani J;  
 XX WPI; 2000-422871/36.  
 DR XX  
 XX XX use of linear peptides as vectors for active ingredients, useful for  
 XX PT diagnosis and treatment of central nervous system diseases, can  
 XX PT transport agents passively across the blood-brain barrier.  
 XX PT  
 XX XX Example I; Page 13; 54pp; French.  
 PS XX  
 XX XX The invention relates to the use of linear peptides, coupled to an active  
 CC agent, to prepare a composition able to cross the blood-brain barrier  
 CC for diagnosis or treatment of disorders localised in the central nervous  
 CC system. The linear peptide preferably has the formula: (a) X1-X16;  
 CC (b); BXXXXXXXBXXXXXB; or (c) BXXXXXXXBXXXXXB, where: each of X1-X16  
 CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
 CC be Trp; each B is aa containing a side chain that includes a basic group;  
 CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
 CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
 CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
 CC peptides able to cross the BBB include protegrins, Antennapedia,  
 CC tachypleins, transportan, etc. Of these several families have cytolytic  
 CC effects and are termed peptide antibiotics. (i) peptides with alpha-helices,  
 CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
 CC beta-sheets, e.g. protegrin, tachypleins, defensins; (iii) peptides  
 CC with no major structure but containing bends due to the presence of  
 CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
 CC fall into the peptide antibiotic categories defined above: (a)-peptides  
 CC are based on the Antennapedia family peptides; (b)-peptides are based on  
 CC protegrins; and (c)-peptides are based on tachypleins. This sequence  
 CC represents a synthetic linear peptide designed on peptides able to cross  
 CC the BBB and is conjugated to a doxorubicin molecule by a succinate  
 CC linker. The peptide may also be linked to a benzylpenicillin molecule  
 CC by a glycoamide linker.  
 CC Conjugates of the linear peptides and the active agent are particularly  
 CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
 CC Parkinson's diseases, depression, pain and meningitis, but also for  
 CC studying drug behaviour in BBB models.  
 XX XX  
 SQ Sequence 18 AA;

Query Match 74.6%; Score 53; DB 21; Length 18;  
 Best Local Similarity 91.7%; Pred. No. 0.004;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RLSYSRRRRFSVS 13  
 Db 4 rlsysrrrrfst 15

RESULT 7  
 ID AAY93179 standard; peptide; 18 AA.  
 XX AC AAY93179;  
 XX AC AAY93179;  
 DT 06-DEC-2000 (first entry)  
 XX Protegrin-like peptide antibiotic Dal-SynB1.

XX Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
 KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
 KW Antennapedia; tachyplein; peptide antibiotic; Alzheimer's disease;  
 KW cancer; Parkinson's disease; depression; pain; meningitis; dalargin.  
 XX Synthetic.

XX Key Location/Qualifiers  
 FT Cross-links 1  
 FT /note= "cross-links to a molecule of dalargin via  
 FT a disulphide linker"

XX WO200032236-A1.

XX 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02938.

XX 30-NOV-1998; 98FR-0015074.

XX (SYNT-) SYNT:EM SA.

XX Clair P, Kaczorek M, Tamsamani J;

XX WPI; 2000-422871/36.

XX use of linear peptides as vectors for active ingredients, useful for  
 XX PT diagnosis and treatment of central nervous system diseases, can  
 XX PT transport agents passively across the blood-brain barrier

XX Example II; Page 20; 54pp; French.

XX The invention relates to the use of linear peptides, coupled to an active  
 CC agent, to prepare a composition able to cross the blood-brain barrier  
 CC for diagnosis or treatment of disorders localised in the central nervous  
 CC system. The linear peptide preferably has the formula: (a) X1-X16;  
 CC (b); BXXXXXXXBXXXXXB; or (c) BXXXXXXXBXXXXXB, where: each of X1-X16  
 CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
 CC be Trp; each B is aa containing a side chain that includes a basic group;  
 CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
 CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
 CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
 CC peptides able to cross the BBB include protegrins, Antennapedia,  
 CC tachypleins, transportan, etc. Of these several families have cytolytic  
 CC effects and are termed peptide antibiotics. (i) peptides with alpha-helices,  
 CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
 CC beta-sheets, e.g. protegrin, tachypleins, defensins; (iii) peptides  
 CC with no major structure but containing bends due to the presence of  
 CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
 CC fall into the peptide antibiotic categories defined above: (a)-peptides  
 CC are based on the Antennapedia family peptides; (b)-peptides are based on  
 CC protegrins; and (c)-peptides are based on tachypleins. This sequence  
 CC represents a synthetic linear peptide designed on peptides able to cross  
 CC the BBB and is conjugated to a dalargin molecule by a disulphide linker.  
 CC Conjugates of the linear peptides and the active agent are particularly  
 CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
 CC Parkinson's diseases, depression, pain and meningitis, but also for  
 CC studying drug behaviour in BBB models.

XX Sequence 18 AA;

Query Match 74.6%; Score 53; DB 21; Length 18;  
 Best Local Similarity 91.7%; Pred. No. 0.004;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RLSYSRRRRFSVS 13  
 Db 4 rlsysrrrrfst 15



Wed Feb 13 07:52:10 2002

08-JUN-2000.

26-NOV-1999; 99WO-FR02939.

30-NOV-1998; 98FR-0015073.

(SYNT-) SYNT:EM SA.

Temsamani J, Kaczorek M, Colin De Verdiere A;

WPI; 2000-412166/35.

New composition useful for cancer treatment and prevention, contains anticancer agent and peptide vector that transports agent into cells

Disclosure: Page 8; 34pp; French.

The specification describes a pharmaceutical composition, which comprises at least one anticancer agent associated with at least one peptide that can transport it into cancer cells and which inhibits development of resistance to the anticancer agent. By using the peptide as a vector for delivery of the anticancer agent, mechanisms that cause cancer cells to become resistant to the agent, particularly the P-glycoprotein pump, are avoided. Also, peptides are easily produced by chemical synthesis, can be coupled easily to the agent, cross mammalian cell membranes rapidly by a passive mechanism (no receptors required), and are non-toxic and non-lytic. The compositions are used to treat cancer. The present sequence represents a peptide which may be linked to the anticancer agents of the invention.

Sequence 10 AA;

Query Match 70.4%; Score 50; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0074;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10

Db 1 RRLSYRRRF 10

RESULT 12

AAW36264

ID AAW36264 standard; peptide; 14 AA.

XX AAW36264;

DT 13-FEB-1998 (first entry)

DE Antimicrobial proteogrin peptide PC34 (64).

Antimicrobial proteogrin; broad spectrum; gram-positive; bacterium; gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus; retrovirus; HIV; human immunodeficiency virus; preservation; disinfection; prophylaxis; treatment; infection; disease; conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD; Helicobacter pylori; sexually transmitted disease; oral mucositis; gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic; respiratory infection; urinary tract infection; MRSA; protozoan; vancomycin resistant Enterococcus; pathogen; multi-drug resistance; penicillin resistant Streptococcus pneumoniae; pig; porcine; methicillin resistant Staphylococcus aureus; systemic candidiasis.

XX Synthetic.

OS OS

XX Sus scrofa.

XX WO9718826-A1.

XX 29-MAY-1997.

XX 22-NOV-1996; 96WO-US18544.

XX

30-NOV-1998; 98FR-0015074.

(SYNT-) SYNT:EM SA.

Clair P, Kaczorek M, Temsamani J;

WPI; 2000-422871/36.

Use of linear peptides as vectors for active ingredients, useful for diagnosis and treatment of central nervous system diseases, can transport agents passively across the blood-brain barrier

Example III; Page 22; 54pp; French.

The invention relates to the use of linear peptides, coupled to an active agent, to prepare a composition able to cross the blood-brain barrier for diagnosis or treatment of disorders localised in the central nervous system. The linear peptide preferably has the formula: (a) X1-X16; (b); BXXBXXXXBXXXXB; or (c) BXXBXXXXBXXXXB, where: each of X1-X16 are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must be Trp; each B is aa containing a side chain that includes a basic group; and each X is an aliphatic or aromatic aa. The linear peptide may be retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment containing at least 5, preferably at least 7 consecutive aa from (a)-(c). Peptides able to cross the BBB include proteogrin, Antennapedia, tachypleins, transportin, etc. Of these several families have cytolytic effects and are termed peptide antibiotics. They fall into 3 main categories based on their structure: (i) peptides with disulphide bond-linked e.g. cecropins and maganins; (ii) peptides with alpha-helices, beta-sheets, e.g. proteogrin, tachypleins, defensins; (iii) peptides with no major structure but containing bends due to the presence of pro residues, e.g. bactericins and PR39. The peptides of the invention fall into the peptide antibiotic categories defined above: (a) peptides are based on the Antennapedia family peptides; (b) peptides are based on proteogrin; and (c) peptides are based on tachypleins. This sequence represents a synthetic linear peptide designed on peptides able to cross the BBB and is conjugated to a doxorubicin molecule by a succinate linker.

Conjugates of the linear peptides and the active agent are particularly used to treat, prevent or diagnose brain cancer, Alzheimer's or Parkinson's diseases, depression, pain and meningitis, but also for studying drug behaviour in BBB models.

Sequence 10 AA;

Query Match 70.4%; Score 50; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0074;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10

Db 1 RRLSYRRRF 10

RESULT 11

ID AAY93618 standard; peptide; 10 AA.

XX AAY93618;

DT 25-SEP-2000 (first entry)

DE Peptide which may be linked to anticancer agents.

Anticancer agent; cancer cell; resistance; P-glycoprotein pump; cancer.

XX Unidentified.

XX WO200032237-A1.

XX



PR 21-NOV-1996; 96US-0752852.  
 PR 22-NOV-1995; 95US-0562346.  
 PR 17-MAY-1996; 96US-0649811.  
 PR 01-AUG-1996; 96US-0690921.  
 XX  
 PA (INTR-) INTRABIOTICS PHARM INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
 XX WPI; 1997-297871/27.  
 XX  
 PT New antimicrobial protegrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)  
 XX  
 PS Claim 23; Page 106; 130pp; English.  
 XX  
 CC The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX  
 SQ Sequence 14 AA;  
 Query Match 64.8%; Score 46; DB 18; Length 14;  
 Best Local Similarity 71.4%; Pred. No. 0.055;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 RRLSYRRRRFSVSV 14  
 ||| | |||| |  
 Db 1 rrlcyrrrrfcv 14  
 RESULT 13  
 AAW36279  
 ID AAW36279 standard; peptide; 16 AA.  
 XX  
 AC AAW36279;  
 XX  
 DT 13-FEB-1998 (first entry)  
 XX  
 DE Antimicrobial protegrin peptide PC34a (79).  
 XX  
 KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
 XX  
 OS Synthetic.

OS Sus scrofa.  
 XX  
 PN WO9718826-A1.  
 XX  
 PD 29-MAY-1997.  
 XX  
 PF 22-NOV-1996; 96WO-US18544.  
 XX  
 PR 21-NOV-1996; 96US-0752852.  
 PR 22-NOV-1995; 95US-0562346.  
 PR 17-MAY-1996; 96US-0649811.  
 PR 01-AUG-1996; 96US-0690921.  
 XX  
 PA (INTR-) INTRABIOTICS PHARM INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
 XX WPI; 1997-297871/27.  
 XX  
 PT New antimicrobial protegrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)  
 XX  
 PS Claim 23; Page 106; 130pp; English.  
 XX  
 CC The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX  
 SQ Sequence 16 AA;  
 Query Match 64.8%; Score 46; DB 18; Length 16;  
 Best Local Similarity 71.4%; Pred. No. 0.064;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 RRLSYRRRRFSVSV 14  
 ||| | |||| |  
 Db 1 rrlcyrrrrfcv 14  
 RESULT 14  
 AAW36440  
 ID AAW36440 standard; peptide; 17 AA.  
 XX  
 AC AAW36440;  
 XX  
 DT 13-FEB-1998 (first entry)  
 XX  
 DE Antimicrobial protegrin peptide (240).  
 XX  
 KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;

XX Helicobacter pylori; sexually transmitted disease; oral mucositis;  
DE gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
XX respiratory infection; urinary tract infection; MRSA; protozoan;  
KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
XX  
OS Synthetic.  
OS Sus scrofa.  
XX  
XX WO9718826-A1.  
PN  
XX  
PD 29-MAY-1997.  
XX  
XX 22-NOV-1996; 96WO-US18544.  
PF  
XX 21-NOV-1996; 96US-0752852.  
PR 22-NOV-1995; 95US-0562346.  
PR 17-MAY-1996; 96US-0649811.  
PR 01-AUG-1996; 96US-0690921.  
XX  
XX (INTR-) INTRABIOTICS PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
PA  
XX Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
XX WPI; 1997-297871/27.  
XX  
XX New antimicrobial protegrin peptide(s) - having activity against  
PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
PT (e.g. HIV)  
PT  
XX  
XX Claim 23; Page 111; 130pp; English.  
PS  
XX The present sequence is an antimicrobial protegrin peptide, which  
XX has a broad spectrum of activity against microbial targets, fungi,  
CC including gram-positive and gram-negative bacteria, yeast, fungi,  
CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
CC It can be used to preserve or disinfect a variety of materials,  
CC including medical equipment, foodstuffs, cosmetics, contact lens  
CC solutions, medicaments or other nutrient containing materials. It  
CC can also be used for the prophylaxis or treatment of microbial  
CC infections or diseases in plants and animals, e.g. conjunctivitis,  
CC keratitis, corneal ulcers, stomach ulcers associated with  
CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
CC sepsis, endocarditis, pneumonia and other respiratory infections,  
CC urinary tract infections, systemic candidiasis and oral mucositis.  
CC It is biostatic or biocidal against clinically relevant pathogens  
CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
CC Enterococcus faecium or faecalis, penicillin resistant  
CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
CC to 1 mg/kg/day, by injection.  
XX  
XX  
SQ Sequence 17 AA;

Query Match 64.8%; Score 46; DB 18; Length 17;  
Best Local Similarity 71.4%; Pred. No. 0.068;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RRLSYRRRFSVSV 14  
||| | ||| | |  
Db 2 rrlcycrrrfccv 15

RESULT 15  
AAW09078  
ID AAW09078 standard; peptide; 17 AA.  
XX  
XX AAW09078;  
XX  
XX 11-AUG-1997 (first entry)  
DT

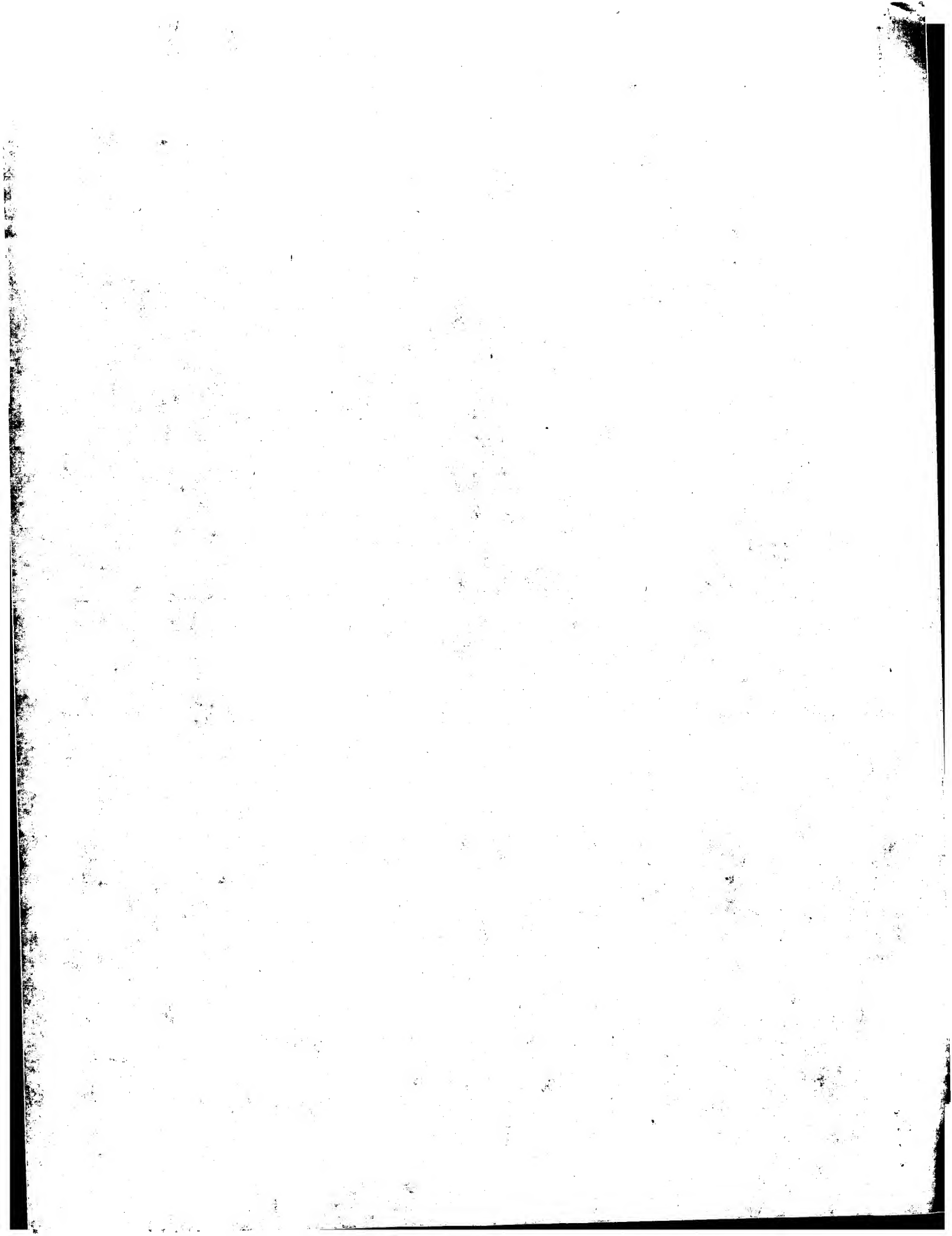
XX Cationic, antimicrobial, virus-neutralising protegrin PC-39.  
DE  
XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
KW Candida albicans; gram-negative bacteria; STD;  
KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;  
KW food.  
XX  
XX Synthetic.  
OS  
XX  
XX Key Location/Qualifiers  
FH Modified-site 1 /note= "Acylated"  
FT Disulfide-bond 6..15  
FT Disulfide-bond 8..13  
FT Modified-site 17 /note= "Amidated"  
FT  
XX WO9637508-A1.  
PN  
XX 28-NOV-1996.  
PD  
XX 24-MAY-1996; 96WO-US07594.  
PF  
XX 07-JUL-1995; 95US-0499523.  
PR 26-MAY-1995; 95US-0451832.  
PR  
XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.  
XX  
XX Harwig SSL, Kokryakov VN, Lehrer RI;  
PI WPI; 1997-033984/03.  
DR  
XX  
XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s)  
PT useful for the treatment of microbial infection, as food  
PT preservatives and in eye care solutions  
XX  
XX Claim 6; Page 63; 106pp; English.  
PS  
XX The present sequence is a specifically claimed example of a peptide,  
CC recombinantly produced, corresponding to the generic formula:  
CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A16-A17-A18)  
CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
CC acid or proline; A17 may be absent or a basic, neutral/polar,  
CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
CC esterified forms, all of which may contain a disulphide bond to give a  
CC cysteine bridge. Peptides of this formula are designated protegrins and  
CC are useful as anti-bacterial, anti-viral and anti-fungal agents in  
CC plants and animals. The protegrins confer resistance to microbial or  
CC viral infection in plants by preventing the growth of a virus or microbe  
CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
CC are particularly useful for the treatment of sexually transmitted  
CC disease caused by microorganisms e.g. Candida albicans, HIV-1, Chlamydia  
CC trachomatis, Treponema pallidum and Neisseria gonorrhoeae. They can also  
CC be used in eye care solutions and as preservatives for food. The  
CC protegrins are more effective under physiological conditions (e.g. in  
CC the presence of serum) than certain antibiotics and are non-toxic to the  
CC cells of higher organisms.  
XX  
XX Sequence 17 AA;

Query Match 64.8%; Score 46; DB 18; Length 17;  
Best Local Similarity 71.4%; Pred. No. 0.068;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 RLSYSRRRFSVSV 15

Db 4 rlcycrrrfcvcv 17

Search completed: February 12, 2002, 12:30:32  
Job time: 365 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:23 ; Search time 106.12 Seconds  
(without alignments)  
3.181 Million cell updates/sec

Title: US-09-485-571-24  
Perfect score: 71  
Sequence: 1 RLSYSRRRFSVSV 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	77.5	18	1	US-08-182-483A-28
2	55	77.5	18	1	US-08-243-879A-27
3	55	77.5	18	1	US-08-499-523-48
4	55	77.5	18	4	US-09-128-345-48
5	46	64.8	14	2	US-08-752-852A-65
6	46	64.8	16	2	US-08-752-852A-80
7	46	64.8	17	2	US-08-752-852A-242
8	45	63.4	18	1	US-08-499-523-63
9	45	63.4	18	1	US-08-499-523-67
10	45	63.4	18	2	US-08-752-852A-86
11	45	63.4	18	2	US-08-752-852A-230
12	45	63.4	18	4	US-09-128-345-63
13	45	63.4	18	4	US-09-128-345-67
14	44	62.0	16	1	US-08-182-483A-20
15	44	62.0	16	1	US-08-243-879A-19
16	44	62.0	16	1	US-08-499-523-40
17	44	62.0	16	1	US-08-499-523-64
18	44	62.0	16	4	US-09-128-345-40
19	44	62.0	16	4	US-09-128-345-64
20	43	60.6	14	2	US-08-752-852A-119
21	43	60.6	18	1	US-08-499-523-53
22	43	60.6	18	1	US-08-499-523-58
23	43	60.6	18	4	US-09-128-345-53
24	43	60.6	18	4	US-09-128-345-58
25	42	59.2	18	1	US-08-499-523-54
26	42	59.2	18	1	US-08-499-523-59
27	42	59.2	18	4	US-09-128-345-54

28	42	59.2	18	4	US-09-128-345-59	Sequence 59, Appl
29	41	57.7	13	2	US-08-752-852A-66	Sequence 66, Appl
30	41	57.7	13	2	US-08-752-852A-240	Sequence 240, App
31	41	57.7	13	3	US-08-752-853-26	Sequence 26, Appl
32	41	57.7	14	2	US-08-752-852A-64	Sequence 64, Appl
33	41	57.7	15	2	US-08-752-852A-62	Sequence 62, Appl
34	41	57.7	15	2	US-08-752-852A-63	Sequence 63, Appl
35	41	57.7	15	2	US-08-752-852A-81	Sequence 81, Appl
36	41	57.7	16	1	US-08-182-483A-5	Sequence 5, Appl1
37	41	57.7	16	1	US-08-182-483A-6	Sequence 6, Appl1
38	41	57.7	16	1	US-08-182-483A-18	Sequence 18, Appl
39	41	57.7	16	1	US-08-243-879A-4	Sequence 4, Appl1
40	41	57.7	16	1	US-08-243-879A-5	Sequence 5, Appl1
41	41	57.7	16	1	US-08-243-879A-17	Sequence 17, Appl
42	41	57.7	16	1	US-08-499-523-21	Sequence 21, Appl
43	41	57.7	16	1	US-08-499-523-22	Sequence 22, Appl
44	41	57.7	16	1	US-08-499-523-38	Sequence 38, Appl
45	41	57.7	16	2	US-08-752-852A-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1  
US-08-182-483A-28  
; Sequence 28, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,483A  
; FILING DATE: 13-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-182-483A-28

Query Match 77.5%; Score 55; DB 1; Length 18;  
Best Local Similarity 84.6%; Pred. No. 0.0024;  
Matches 11, Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 RLSYSRRRFSVSV 14  
I:|:|||||  
Db 4 RLSYSRRRFSVSV 16

us-09-485-571-24.ra1

Wed Feb 13 07:52:11 2002

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-499-523-48

```

```

Query Match 77.5%; Score 55; DB 1; Length 18;
Best Local Similarity 84.6%; Pred. No. 0.0024;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 RLSYSRRRFSVS 14
Db 4 RISFSRRRFSVS 16

```

```

RESULT 4
US-09-128-345-48
; Sequence 48, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-128-345-48

```

```

RESULT 2
US-08-243-879A-27
; Sequence 27, Application US/08243879A
; Patent No. 5708145
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; TITLE OF INVENTION: A NEW PROTEGRIN
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington, DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/243,879A
; FILING DATE: 17-MAY-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-243-879A-27

```

```

Query Match 77.5%; Score 55; DB 1; Length 18;
Best Local Similarity 84.6%; Pred. No. 0.0024;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 RLSYSRRRFSVS 14
Db 4 RISFSRRRFSVS 16

```

```

RESULT 3
US-08-499-523-48
; Sequence 48, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

RESULT 6  
US-08-752-852A-80  
: Sequence 80, Application US/08752852A  
: Patent No. 5994306

```

:
: GENERAL INFORMATION:
:
: APPLICANT: Chang, Conway
: APPLICANT: Gu, Chee-liang
: APPLICANT: Chen, Jie
: APPLICANT: Steinberg, Deborah
: APPLICANT: Lehrer, Robert
: APPLICANT: Harwig, Sylvia
: TITLE OF INVENTION: FINE-TUNE
:
: NUMBER OF SEQUENCES: 242
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the
: CITY: New York
: STATE: NY
:
:

```

us-09-485-571-24.ra1

Wed Feb 13 07:52:11 2002

```

;
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752.852A
; FILING DATE: 21-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141
; INFORMATION FOR SEQ ID NO: 242:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-752-852A-242

Query Match 64.8%; Score 46; DB 2; Length 17;
Best Local Similarity 71.4%; Pred. No. 0.073;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RRLSYRRRFSVSV 14
   ||| ||||| |
Db 2 RRLCYRRRFCV 15

RESULT 8
US-08-499-523-63
; Sequence 63, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-08-499-523-67

;
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752.852A
; FILING DATE: 21-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141
; INFORMATION FOR SEQ ID NO: 242:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-752-852A-242

Query Match 64.8%; Score 46; DB 2; Length 17;
Best Local Similarity 71.4%; Pred. No. 0.073;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RRLSYRRRFSVSV 14
   ||| ||||| |
Db 2 RRLCYRRRFCV 15

RESULT 8
US-08-499-523-63
; Sequence 63, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-08-499-523-67

;
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752.852A
; FILING DATE: 21-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141
; INFORMATION FOR SEQ ID NO: 242:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-752-852A-242

Query Match 63.4%; Score 45; DB 1; Length 18;
Best Local Similarity 69.2%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 RLSYSRRRFSVSV 14
   || | |||| |
Db 4 RLXYRRRFXV 16

RESULT 9
US-08-499-523-67
; Sequence 67, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-08-499-523-67
```



Query Match 63.4%; Score 45; DB 1; Length 18;  
Best Local Similarity 69.2%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 RLSYRRRFSVSV 14  
||| ||||| |  
Db 4 RLXYRRRFXVXV 16

## RESULT 10

US-08-752-852A-86  
; Sequence 86, Application US/08752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrner, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141

; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-752-852A-86

Query Match 63.4%; Score 45; DB 2; Length 18;  
Best Local Similarity 69.2%; Pred. No. 0.11;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYRRRFSVSV 14  
||| ||||| |  
Db 4 RLAYRRRFXVAV 16

## RESULT 11

US-08-752-852A-230  
; Sequence 230, Application US/08752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrner, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrner, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141

; INFORMATION FOR SEQ ID NO: 230:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-752-852A-230

Query Match 63.4%; Score 45; DB 2; Length 18;  
Best Local Similarity 69.2%; Pred. No. 0.11;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYRRRFSVSV 14  
||| ||||| |  
Db 4 RLXYRRRFXVXV 16

## RESULT 12

US-09-128-345-63  
; Sequence 63, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KORZYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

us-09-485-571-24.ra1

wed Feb 13 07:52:11 2002

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-67

```

```

Query Match 63.4%; Score 45; DB 4; Length 18;
Best Local Similarity 69.2%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 RLSYSRRRFSVSV 14
   ||| |||||
Db 4 RLXYRRRFXV 16

```

```

RESULT 14
US-08-182-483A-20
; Sequence 20, Application US/08182483A
; Patent No. 5693486
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: KORYAKOV, VLADIMIR N.
; APPLICANT: HARWIG, SYLVIA S.L.
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182.483A
; FILING DATE: 13-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-182-483A-20

```

```

Query Match 62.0%; Score 44; DB 1; Length 16;
Best Local Similarity 69.2%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 RLSYSRRRFSVSV 14
   ||| |||||
Db 4 RLGYRRRFGVCV 16

```

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-63

```

```

Query Match 63.4%; Score 45; DB 4; Length 18;
Best Local Similarity 69.2%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 RLSYSRRRFSVSV 14
   ||| |||||
Db 4 RLXYRRRFXV 16

```

```

RESULT 13
US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:

```

RESULT 15  
US-08-243-879A-19  
: Sequence 19, Application US/08243879A  
: Patent No. 5708145  
: GENERAL INFORMATION:  
: APPLICANT: LEHRER, ROBERT I.  
: APPLICANT: HARWIG, SYLVIA S.L.  
: TITLE OF INVENTION: A NEW PROTEGRIN  
: NUMBER OF SEQUENCES: 42  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: MORRISON & FOERSTER  
: STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
: CITY: Washington, DC  
: COUNTRY: USA  
: ZIP: 20006-1812  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/243,879A  
: FILING DATE: 17-MAY-1994  
: CLASSIFICATION: 530  
: ATTORNEY/AGENT INFORMATION:  
: NAME: MURASHIGE, KATE H.  
: REGISTRATION NUMBER: 29,959  
: REFERENCE/DOCKET NUMBER: 2000-0540.22  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (202) 887-1500  
: TELEFAX: (202) 887-0763  
: TELEX: 90-4030  
: INFORMATION FOR SEQ ID NO: 19:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 16 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
US-08-243-879A-19

Query Match 62.0%; Score 44; DB 1; Length 16;  
Best Local Similarity 69.2%; Pred. No. 0.15;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 RLSYSRRRRFSVSV 14  
Db 4 RLGYSRRRRFGVCV 16

Search completed: February 12, 2002, 12:32:23  
Job time: 451 sec





Wed Feb 13 07:52:12 2002

us-09-485-571-24.rpr

Db 65 RRRYSRRY 74

|| :||||| :||

Db 50 RRGYSRRSFTVS 62

RESULT 3

S56116  
spermatid-specific protein T1 - longfin squid  
N:Alternate names: arginine-rich protamine; testis-specific protein T2  
C:Species: Loligo pealeii (longfin squid)  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 07-May-1999  
C:Accession: S56116  
R:Wouters-Tyrou, D.; Martin-Ponthieu, A.; Ledoux-Andula, N.; Kouach, M.; Jaquinod, I  
Biochem. J. 309, 529-534, 1995  
A:Title: Squid spermatogenesis: molecular characterization of testis-specific pro-prc  
A:Reference number: S56116; MUID:95351983  
A:Accession: S56116  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-79 <WOU>  
C:Superfamily: sperm histone

Query Match 57.7%; Score 41; DB 2; Length 79;  
Best Local Similarity 80.0%; Pred. No. 2.8;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLYSRRRF 10  
|| :||||| :||

Db 66 RRRYSRRY 75

RESULT 6

S57607  
protegrin 1 precursor - pig  
N:Alternate names: neutrophil peptide 1  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S66284; S45712; S36820; S34585; S57607  
R:Zhao, C.; Ganz, T.; Lehrer, R.I.  
FEBS Lett. 368, 197-202, 1995  
A:Title: The structure of porcine neutrophil genes.  
A:Reference number: S66283; MUID:95354835  
A:Accession: S66284  
A:Molecule type: DNA  
A:Residues: 1-149 <ZHA>  
A:Cross-references: EMBL:X84094; NID:9887642; PIDN:CAA58890.1; PID:9887643  
R:Zhao, C.; Liu, L.; Lehrer, R.I.  
FEBS Lett. 346, 285-288, 1994  
A:Title: Identification of a new member of the protegrin family by cDNA cloning.  
A:Reference number: S45712; MUID:94283613  
A:Accession: S45712  
A:Molecule type: mRNA  
A:Residues: 1-149 <ZHA>  
A:Cross-references: GB:X79868; NID:9603035; PIDN:CAA56251.1; PID:9603036  
R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; E  
FEBS Lett. 330, 339-342, 1993  
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Sequ  
A:Reference number: S36820; MUID:93387466  
A:Accession: S36820  
A:Molecule type: protein  
A:Residues: 131-148 <MR>  
R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M  
FEBS Lett. 327, 231-236, 1993  
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of corti  
A:Reference number: S34585; MUID:93327946  
A:Accession: S34585  
A:Molecule type: protein  
A:Residues: 131-148 <KOK>  
C:Genetics:  
A:Gene: NPGL  
A:Introns: 66/3; 102/3; 126/3  
C:Superfamily: cathelin; cystatin homology  
C:Keywords: amidated carboxyl end; antibacterial; neutrophil  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:22-129/Domain: cystatin homology <CYS>

Query Match 57.7%; Score 41; DB 2; Length 78;  
Best Local Similarity 80.0%; Pred. No. 2.8;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLYSRRRF 10  
|| :||||| :||

RESULT 4

A40973  
spermatid-specific protein T1 precursor - common cuttlefish  
N:Alternate names: arginine-rich protamine; testis-specific protein T1  
C:Species: Sepia officinalis (common cuttlefish)  
C:Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Sep-1998  
C:Accession: A40973; S14085  
R:Wouters-Tyrou, D.; Chartier-Harlin, M.C.; Martin-Ponthieu, A.; Boutillon, C.; Van Dor  
J. Biol. Chem. 266, 17388-17395, 1991  
A:Title: Cuttlefish spermatid-specific protein T. Molecular characterization of two vari  
A:Reference number: A40973; MUID:91153298  
A:Accession: A40973  
A:Molecule type: protein  
A:Residues: 1-78 <WOU>  
R:Martin-Ponthieu, A.; Wouters-Tyrou, D.; Belaiche, D.; Sautiere, P.; Schindler, P.; van  
Eur. J. Biochem. 195, 611-619, 1991  
A:Title: Cuttlefish sperm protamines. 1. Amino acid sequences of two distinct variants.  
A:Reference number: S14085; MUID:91153298  
A:Accession: S14085  
A:Molecule type: protein  
A:Residues: 22-78 <MAR>  
C:Superfamily: sperm histone  
C:Keywords: DNA binding; nucleus; phosphoprotein; spermatogenesis  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-78/Product: protamine variant Sp1 #status experimental <MAT>

Query Match 57.7%; Score 41; DB 2; Length 78;  
Best Local Similarity 80.0%; Pred. No. 2.8;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLYSRRRF 10  
|| :||||| :||

F:30-130/Domain: propeptide #status predicted <PRO>  
 F:131-148/Product: proteogrin 1 #status experimental <MAT>  
 F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from following gl

Query Match 57.7%; Score 41; DB 2; Length 149;  
 Best Local Similarity 59.1%; Pred. No. 5.1;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 RLVSYSRRRFSVSV 14

Db 134 RLVCRRRRCV 146

RESULT 7

A60234  
 IGA Fc receptor precursor - Streptococcus agalactiae (strain SB35)

N:Alternate names: IGA-binding protein; protein Bac  
 N:Contains: beta antigen

C:Species: Streptococcus agalactiae

C:Date: 08-Dec-1992 #sequence\_revision 08-Dec-1992 #text\_change 26-Aug-1999

C:Accession: A60234; S14595; A60230

R:Heden, L.O.; Frithz, E.; Lindahl, G.

Eur. J. Immunol. 21, 1481-1490, 1991

A:Title: Molecular characterization of an IGA receptor from group B streptococci: sequen  
 ents with IGA-binding capacity.

A:Reference number: A60234; MUID:91257158

A:Accession: A60234

A:Molecule type: DNA

A:Residues: 1-1134 <HE2>

A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521

A:Note: the source is designated as Streptococcus strain SB35

R:Lindahl, G.; Akerstroem, B.; Vaerman, J.P.; Stenberg, L.

Eur. J. Immunol. 20, 2241-2247, 1990

A:Title: Characterization of an IGA receptor from group B streptococci: specificity for  
 fragments.

A:Reference number: S14595

A:Accession: S14595

A:Molecule type: DNA

A:Residues: 1-1134 <HE2>

A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521

A:Note: the source is designated as Streptococcus agalactiae

R:Lindahl, G.; Akerstroem, B.; Vaerman, J.P.; Stenberg, L.

Eur. J. Immunol. 20, 2241-2247, 1990

A:Title: Characterization of an IGA receptor from group B streptococci: specificity for

A:Reference number: A60230; MUID:91055597

A:Accession: A60230

A:Molecule type: protein

A:Residues: 'X',39-48,'X',50-52,'X',54-56 <LIN>

C:Superfamily: IGA Fc receptor

C:Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein

F:1-37/Domain: signal sequence #status predicted <SIG>

F:38-1134/Product: IGA Fc receptor #status predicted <MAT>

F:199-438/Domain: IGA binding #status experimental

F:439-826/Domain: IGA binding #status predicted <IGAL>

F:827-915/Region: IGA binding #status predicted <IGA2>

F:916-1101/Domain: cell wall-spanning #status predicted <CWS>

F:1102-1129/Domain: transmembrane #status predicted <TMM>

Query Match 57.7%; Score 41; DB 2; Length 1134;

Best Local Similarity 57.1%; Pred. No. 33;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RLVSYSRRRFSVSV 14

Db 8 RKMYSIRKFSVGV 21

RESULT 8

FCSOAG

IGA Fc receptor precursor - Streptococcus agalactiae

N:Alternate names: beta antigen

C:Species: Streptococcus agalactiae

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999

C:Accession: S15330; S20240; S17038

R:Jerlstroem, P.G.; Chhatwal, G.S.; Timmis, K.N.

Mol. Microbiol. 5, 843-849, 1991

A:Title: the IGA-binding beta antigen of the c protein complex of Group B streptococ

A:Reference number: S15330; MUID:91312121

A:Accession: S15330

A:Molecule type: DNA

A:Residues: 1-1164 <JER1>

A:Cross-references: EMBL:X59771

A:Accession: S20240

A:Molecule type: protein

A:Residues: 38-48 <JE2>

R:Jerlstroem, P.G.

submitted to the EMBL Data Library, August 1991

A:Reference number: S17038

A:Accession: S17038

A:Molecule type: DNA

A:Residues: 1-914,'E',916-1164 <JE3>

A:Cross-references: EMBL:X59771; NID:g46522; PIDN:CAA42442.1; PID:g46523

C:Superfamily: IGA Fc receptor

C:Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protei

F:1-37/Domain: signal sequence #status predicted <SIG>

F:38-1164/Product: IGA Fc receptor #status experimental <MAT>

F:199-438/Domain: IGA binding #status predicted <IGAL>

F:439-826/Domain: IGA binding #status predicted <IGA2>

F:827-945/Region: proline-rich repeats

F:946-1131/Domain: cell wall-spanning #status predicted <CWS>

F:1132-1159/Domain: transmembrane #status predicted <TMM>

Query Match 57.7%; Score 41; DB 1; Length 1164;

Best Local Similarity 57.1%; Pred. No. 34;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RLVSYSRRRFSVSV 14

Db 8 RKMYSIRKFSVGV 21

RESULT 9

DB1096

hypothetical protein NMB1317 [imported] - Neisseria meningitidis

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: DB1096

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty,

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58

A:Reference number: A81000; MUID:20175755

A:Accession: DB1096

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-134 <JET>

A:Cross-references: GB:AE002480; GB:AE002098; NID:g7226555; PIDN:AAF41692.1; PID:g7

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB1317

Query Match 56.3%; Score 40; DB 2; Length 134;

Best Local Similarity 46.7%; Pred. No. 6.9;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 RLVSYSRRRFSVSV 15

Db 34 KTLSTNLSRPFKISIR 48

```

Query Match      56.3%; Score 40; DB 2; Length 249;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 RRLSYRRRFRFSVSVR 15
    ||| ||||| |||
Db 49 RRLSYRRRFRFSVSVR 63

RESULT 12
F69309
A:Protein: protein Phnp (phnp) homolog - Archaeoglobus fulgidus
A:Species: Archaeoglobus fulgidus
A:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
A:Accession: F69309
A:Residues: 1-147 <STO>
A:Title: A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide with a cat
A:Reference number: JN0900; MUID:94071898
A:Accession: JN0900
A:Molecule type: mRNA
A:Residues: 1-147 <STO>
A:Cross-references: GB:L24745; NID:9431435; PID:9431436
A:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Egorov,
FEBs Lett. 330, 339-342, 1993
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Sequence
A:Reference number: S36820; MUID:93387466
A:Accession: S36822
A:Molecule type: protein
A:Residues: 131-146 <MIR>
A:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.; Sh
FEBs Lett. 327, 231-236, 1993
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of corticost
A:Reference number: S34585; MUID:93327946
A:Accession: S34586
A:Molecule type: protein
A:Residues: 131-146 <KOK>
A:Comment: This protein plays a role in the nonoxidative antimicrobial defense mechanism
C:Superfamily: cathelin; cystatin homolog
C:Keywords: amidated carboxyl end; antibacterial; neutrophil
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homolog <CYS>
F:131-146/Product: protegrin 2 #status experimental <MAT>
F:146/Modified site: amidated carboxyl end (Val) (amide in mature form from following gl

Query Match      56.3%; Score 40; DB 2; Length 147;
Best Local Similarity 61.5%; Pred. No. 7.5;
Matches 8; Conservative 1; Mismatches 0; Gaps 0;

Oy 2 RLSYRRRFRFSVSV 14
    ||| ||||| |||
Db 134 RLCYRRRFRFCICV 146

RESULT 11
B86273
A:Hypothetical protein AAD39285.1 [imported] - Arabidopsis thaliana
A:Species: Arabidopsis thaliana (mouse-ear cress)
A:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A:Accession: B86273
A:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B86273
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <STO>
A:Cross-references: GB:AE005172; NID:95080775; PIDN:AAD39285.1; GSPDB:GNC0141
C:Genetics:
A:Map position: 1

Query Match      56.3%; Score 40; DB 2; Length 254;
Best Local Similarity 53.3%; Pred. No. 12;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 RRLSYRRRFRFSVSVR 15
    ||| ||||| |||
Db 28 RRGWRRFRFSVMVQ 42

RESULT 13
T47575
A:Hypothetical protein F24B22.120 - Arabidopsis thaliana
A:Species: Arabidopsis thaliana (mouse-ear cress)
A:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 17-Nov-2000
A:Accession: T47575
A:Blocker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223016
A:Accession: T47575
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-434 <BLO>
A:Cross-references: EMBL:AL32957
A:Experimental source: cultivar Columbia; BAC clone F24B22
C:Genetics:
A:Map position: 3
A:Introns: 284/3; 331/3
A:Note: F24B22.120
A:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

Query Match      56.3%; Score 40; DB 2; Length 434;
Best Local Similarity 60.0%; Pred. No. 20;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 RRLSYRRRFRFSVSVR 15
    ||| ||||| |||
Db 56 RRDYSRRFRFSVSVR 70

RESULT 14
S56117
A:spmatid-specific protein T2 precursor - longfin squid

```



N;Alternate names: sperm protamin SP  
C:Species: Loligo pealeii (longfin squid)  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 07-May-1999  
C:Accession: S56117  
R:Wouters-Tyrou, D.; Martin-Ponthieu, A.; Ledoux-Andula, N.; Kouach, M.; Jaquinod, M.; S  
Biochem. J. 309, 529-534, 1995  
A:Title: Squid spermiogenesis: molecular characterization of testis-specific pro-protami  
A:Reference-number: S56116; MUID:95951983  
A:Accession: S56117  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-118 <WOU>

Query Match 53.5%; Score 38; DB 2; Length 118;  
Best Local Similarity 88.9%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYSRRR 9  
|||  
Db 65 RRRSYSRRR 73

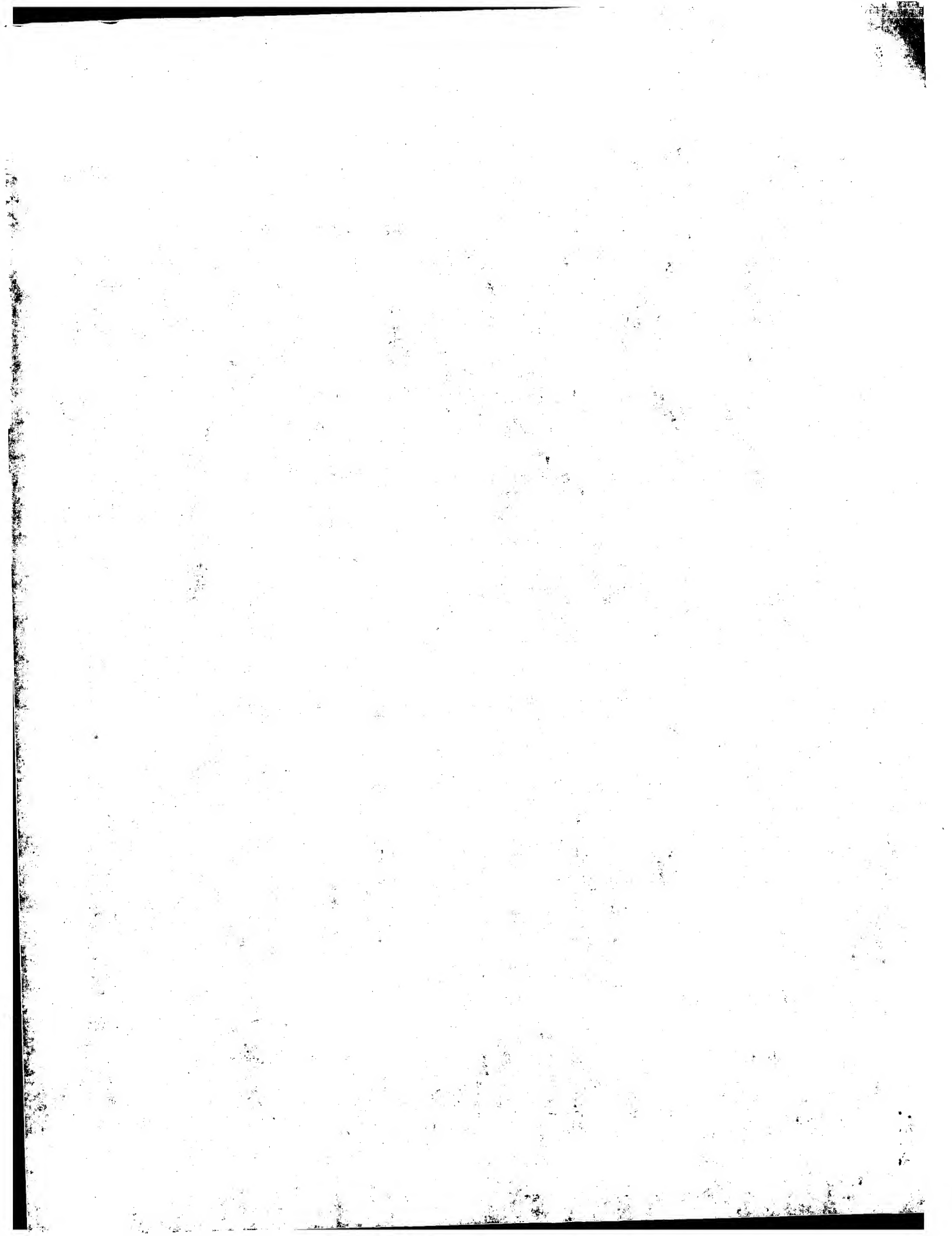
## RESULT 15

S63648  
H+-transporting ATP synthase (EC 3.6.1.34) protein 6 - Allomyces macrogynus mitochondri  
C:Species: mitochondrion Allomyces macrogynus  
C:Date: 19-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 22-Jun-1999  
C:Accession: S63648  
R:Paquin, B.; Lang, B.F.  
J. Mol. Biol. 255, 688-701, 1996  
A:Title: The mitochondrial DNA of Allomyces macrogynus: the complete genomic sequence fr  
A:Reference-number: S63635; MUID:96226032  
A:Accession: S63648  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <PAQ>  
A:Cross-references: EMBL:U41288; NID:g1236403; PIDN:AAC49231.1; PID:g1236414  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995  
A>Note: the C-terminal part of this protein originates from a foreign atp6 gene  
R:Paquin, B.; Laforest, M.J.; Lang, B.F.  
Proc. Natl. Acad. Sci. U.S.A. 91, 11807-11810, 1994  
A:Title: Interspecific transfer of mitochondrial genes in fungi and creation of a homolo  
A:Reference-number: S77735; MUID:95083591  
A:Contents: annotation; horizontal gene transfer  
C:Genetics:  
A:Gene: atp6  
A:Genome: mitochondrion  
C:Superfamily: H+-transporting ATP synthase protein 6  
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; ox

Query Match 53.5%; Score 38; DB 2; Length 262;  
Best Local Similarity 53.8%; Pred. No. 28;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 LSYSRRRFSVSR 15  
:||| | |:::|  
Db 172 ISYSARAFSLALR 184

Search completed: February 12, 2002, 12:34:40  
Job time: 558 sec



Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	41	57.7	59	1	HSP1_MACRU	P42142 macroopus ru
2	41	57.7	60	1	HSP1_CAERU	P42131 caenolestes
3	41	57.7	60	1	HSP1_DASVI	P42135 dasyurus vi
4	41	57.7	60	1	HSP1_MACAG	P42137 macropus ag
5	41	57.7	60	1	HSP1_MACGI	P42139 macropus gi
6	41	57.7	61	1	HSP1_ANTLA	O18745 antechinus
7	41	57.7	61	1	HSP1_ANTSW	P42130 antechinus
8	41	57.7	61	1	HSP1_MACEU	P42138 macropus eu
9	41	57.7	61	1	HSP1_MACRG	P42141 macropus ru
10	41	57.7	61	1	HSP1_PARBI	O18768 parantechin
11	41	57.7	61	1	HSP1_SARHA	P42151 sarcophilus
12	41	57.7	61	1	HSP1_TRIVU	P42152 trichosurus
13	41	57.7	62	1	HSP1_DASRO	P42134 dasykaluta
14	41	57.7	62	1	HSP1_MURLO	P42140 murexia lon
15	41	57.7	63	1	HSP1_ANTST	P42129 antechinus
16	41	57.7	77	1	PR12_SEPOF	P80002 sepia offic
17	41	57.7	78	1	PR11_SEPOF	P80001 sepia offic
18	41	57.7	149	1	PG1_PIG	P32194 sus scrofa
19	41	57.7	1164	1	BAG_STRAG	P27951 streptococ
20	40	56.3	147	1	PG2_PIG	P32195 sus scrofa
21	38	53.5	262	1	ATP6_ALLAR	P50363 allomyces a
22	38	53.5	265	1	ATP6_ALLMA	P50364 allomyces m
23	38	53.5	295	1	GEM_MOUSE	P55041 mus musculus
24	38	53.5	499	1	UDPH_SCHPO	O59819 schizosacch
25	37	52.1	241	1	ATP6_RHORU	P15012 rhodospiril
26	36.5	51.4	339	1	LPXK_XYLFA	Q9pee6 xyella fas
27	36	50.7	113	1	RS6_SYNY3	P73636 synechocyst
28	36	50.7	149	1	PG3_PIG	P32196 sus scrofa
29	36	50.7	612	1	YMY8_YEAST	Q03153 saccharomyc
30	36	50.7	624	1	NIFA_AZOLI	P54929 azospirillu
31	36	50.7	625	1	NIFA_AZOB	P30667 azospirillu
32	36	50.7	698	1	CVAB_ECOLI	P22520 escherichia
33	36	50.7	937	1	ROD1_YEAST	Q02805 saccharomyc

P42131;  
01-NOV-1995 (Rel. 32, Created)  
01-NOV-1995 (Rel. 32, Last sequence update)  
01-NOV-1997 (Rel. 35, Last annotation update)  
SPERM PROTAMINE P1.  
PRM1.  
Caenolestes fuliginosus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Metatheria; Paucituberculata; Caenolestidae; Caenolestes.  
NCBI\_TaxID=37696;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Sperm;  
MEDLINE=95215351; PubMed=7700877;  
Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;  
"Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).  
-!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
-!- SUBCELLULAR LOCATION: NUCLEAR.  
-!- TISSUE SPECIFICITY: TESTIS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: L35332; AAA74598.1; -  
InterPro: IPRO00221; Protamine.P1.  
Pfam: PF00260; protamine.P1; 1.  
PROSITE: PS00048; PROTAMINE\_P1; 1.  
DR: Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
KW: Testis; DNA condensation; Nuclear protein.  
FT: 0 BY SIMILARITY  
INIT\_MET 0  
SEQUENCE 60 AA; 8514 MW; 7630E63AD33A9B05 CRC64;  
SQ

```

Query Match          57.7%; Score 41; DB 1; Length 60;
Best Local Similarity 72.7%; Pred. No. 0.4;
Matches      8; Conservative      1; Mismatches      2; Indels      0; Gaps      0;

```

```

1 RRLSYRRRFS 11
  || |||||:|
43 RRGYSRRRYS 53

```

RESULT	3	STANDARD;	PRT;	60 AA.
ID	HSPI_DASVI			
CD	P42135; P42133;			
DC	01-NOV-1995	(Rel. 32, Created)		
DD	01-NOV-1995	(Rel. 32, Last sequence update)		
DE	15-JUL-1999	(Rel. 38, Last annotation update)		
DF	SP5PM PROTAINE P1.			

PRM1.  
Dasyurus viverrinus (Southeastern quoll), and  
Dasyurus hallucatus (Satanellus/northern quoll).  
OS Ooscytus: Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasyurus.  
OC NCBI\_TaxID=9279, 9280;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Sperm;  
RC MEDLINE=95215351; PubMed=7700877;  
RX Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;  
RA "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
RT Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
RT -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY CONTACT  
CC

—

SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
-!- SUBCELLULAR LOCATION: NUCLEAR.  
-!- TISSUE SPECIFICITY: TESTIS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; L35340; AAA74599.1; -
DR	EMBL; L35341; AAA56795.1; -
DR	InterPro; IPR000221; Protamine_pl.
DR	Pfam; PF00260; Protamine_P1; 1.
DR	PROSITE; PS00048; PROTAMINE_P1; 1.
DR	Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW	Testis; DNA condensation; Nuclear protein.
KW	Testis; DNA condensation; Nuclear protein.
INIT MET	0 BY SIMILARITY.
FO	SEQUENCE 60 AA; 8246 MW; 615D3D85E7123025 CRC64;

```

Query Match          57.7%; Score 41; DB 1; Length 60;
Best Local Similarity 72.7%; Pred. No. 0.4;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 RRLSYSTRRRES 11
      ||| |||||:|
Db     44 RRRGYSRRRYS 54

```

RESULT 4  
HSP1\_MACAG  
STANDARD  
PRT: 60 AA.

AC P42137;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Sperm Protamine Pl.

FRUL.  
GN  
OS  
OC

OC Mammalia; Metachiro  
OC NCBI TaxID=9313.

```

UX NCBI_IAXID-5515,
RN [1]
PR SEQUENCE FROM N.A.

```

```
RC TISSUE=Sperm;
SEQUENCE FROM NAME;
RP
```

TABLE 1  
RX  
MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westerman M., Winkler R.J., Dixon G.H.,  
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
250.7-1.7/1995)

Proc. R. Soc. Lond., B, Biol. Sci. 239:7-14 (1995).

CC DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THE  
SPERM DVA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC SPERM DNA INTO A HIGHLY CONDENSED NUCLEAR LOCATION: NUCLEAR: -1-

CC - ! - SUBCELLULAR LOCATION: NUCLEUS  
CC - ! - TISSUE SPECIFICITY: TESTIS.

38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

CC This SWISS-PROT entry is copyright

CC between the Swiss Institute of

the European Bioinformatics Institute

CC use by non-profit institutions is modified and this statement is modified and this statement is

entitles requires a license agreement

CC or send an email to [license@ish.ch](mailto:license@ish.ch)

22

DR EMBL; L35451; AAA74615.1; -;  
-; F00000231; Protamine

DR InterPro; IPR000221; Pfocamine.  
DR Pfam; PF00260; protamine P1: 1

BR PROTAM; PE00290; PROCAINE-1;  
BR PROTAM; PS00048: PROTAMINE\_PL  
BR PROSITE: PS00048: PROTAMINE\_PL

Chromosomal protein; Nucleosome

KW  
KW

Testis; DNA condensation; Nucl

FT	INIT_MET	0	0	B
----	----------	---	---	---

SO SEQUENCE 60 AA; 8338 MW; 96255C818921EB85 CRC64;

Query Match 57.7%; Score 41; DB 1; Length 60;  
Best Local Similarity 72.7%; Pred. No. 0.4;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRLSYRRRFS 11  
II IIIII:|  
DB 43 RRGYSRRRYS 53

RESULT 5

HSPI\_MACGI STANDARD; PRT; 60 AA.  
AC P42130;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.

OS Macropus giganteus (Eastern gray kangaroo)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.  
OX NCBI\_TaxID=9317;  
RN [1]

SEQUENCE FROM N.A.

TISSUE-Sperm;

RA MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;  
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes."; Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).

CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- TISSUE SPECIFICITY: TESTIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: L35333; AAA74604.1; -

DR InterPro: IPR000221; Protamine\_P1.

DR Pfam: PF00260; protamine\_P1; 1.

DR PROSITE: PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;

KW Testis; DNA condensation; Nuclear protein.

FT INIT\_MET 0 BY SIMILARITY.

SO SEQUENCE 60 AA; 8415 MW; 1DC25C80C490BC90 CRC64;

Query Match 57.7%; Score 41; DB 1; Length 60;  
Best Local Similarity 72.7%; Pred. No. 0.4;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRLSYRRRFS 11  
II IIIII:|  
DB 44 RRGYSRRRYS 54

RESULT 6

HSPI\_ANTLA STANDARD; PRT; 61 AA.  
AC O18745;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE SPERM PROTAMINE P1.

OS Antechinus swainsonii, Phascosorex dorsalis,

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.

OX NCBI\_TaxID=9284, 9295, 32551, 32545, 63143, 32546;

RN [1]

SEQUENCE FROM N.A.

SPECIES=A.swainsonii, and P.dorsalis; TISSUE=Sperm;

MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;

"Molecular phylogeny and evolution of marsupial protamine P1 genes."; Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).

GN PRM1.

OS Antechinomya laniger.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinomya.

OX NCBI\_TaxID=60701;

RN [1]

SEQUENCE FROM N.A.

MEDLINE=97446280; PubMed=9299228;

RA Krajewski C., Blacket M., Buckley L., Westernman M.;

RT "A multigene assessment of phylogenetic relationships within the dasyurid marsupial subfamily Sminthopsinae."; Mol. Phylogenet. Evol. 8:236-248(1997).

RL Mol. Phylogenet. Evol. 8:236-248(1997).

CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- TISSUE SPECIFICITY: TESTIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF001587; AAB91377.1; -

DR InterPro: IPR000221; Protamine\_P1.

DR Pfam: PF00260; protamine\_P1; 1.

DR PROSITE: PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;

KW Testis; DNA condensation; Nuclear protein.

FT INIT\_MET 0 BY SIMILARITY.

SO SEQUENCE 61 AA; 8409 MW; E021567627E562B1 CRC64;

Query Match 57.7%; Score 41; DB 1; Length 61;  
Best Local Similarity 72.7%; Pred. No. 0.41;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 RRLSYRRRFS 11  
II IIIII:|  
DB 44 RRGYSRRRYS 54

RESULT 7

HSPI\_ANTSW STANDARD; PRT; 61 AA.

AC P42130; P42146;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE SPERM PROTAMINE P1.

GN PRM1.

OS Antechinus swainsonii, Phascosorex dorsalis,

OC Neophascogale lorentzii (Long-clawed marsupial mouse),

OS Dasyurus albopunctatus (Native cat), and

OS Dasyurus geoffroii (Chuditch/western quoll), and

OS Dasyurus spartacus (Native cat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.

OX NCBI\_TaxID=9284, 9295, 32551, 32545, 63143, 32546;

RN [1]

SEQUENCE FROM N.A.

SPECIES=A.swainsonii, and P.dorsalis; TISSUE=Sperm;

MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;

"Molecular phylogeny and evolution of marsupial protamine P1 genes."; Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).

RN [2]

SEQUENCE FROM N.A.

SPECIES=N.lorentzii, D.albopunctatus, D.geoffroii, and D.spartacus;

RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;

Wed Feb 13 07:52:13 2002

```
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
RT cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35338; AAB95429.1; -
CC EMBL; L35339; AAA74601.1; -
CC EMBL; AF010267; AAB69327.1; -
CC EMBL; AF010272; AAB69302.1; -
CC EMBL; AF010274; AAB69304.1; -
CC EMBL; AF010275; AAB69305.1; -
CC InterPro: IPR000221; Protamine_P1.
CC PROSITE: PS00048; Protamine_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
KW INIT_MET 0 BY SIMILARITY.
FT SEQUENCE 61 AA; 8390 MW; E021472785E71221 CRC64;
SQ
Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 RRLSYRRRRS 11
DB 44 RRGYSRRRYS 54
RESULT 8
HSP1_MACEU STANDARD; PRT; 61 AA.
ID HSP1_MACEU
AC P42138;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1
OS Macropus eugenii (Tamar wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9315;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE-Specific;
RA MEDLINE=95215351; PubMed=7700877;
RL "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35329; AAA74610.1; -
CC EMBL; L35328; AAA74609.1; -
CC InterPro: IPR000221; Protamine_P1.
CC Pfam: PF00260; protamine_P1; 1.
CC PROSITE: PS00048; Protamine_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
KW INIT_MET 0 BY SIMILARITY.
FT SEQUENCE 61 AA; 8571 MW; B822925C80C490BC CRC64;
SQ
Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 RRLSYRRRRS 11
DB 44 RRGYSRRRYS 54
RESULT 8
HSP1_MACEU STANDARD; PRT; 61 AA.
ID HSP1_MACEU
AC P42138;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1
OS Macropus eugenii (Tamar wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9315;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE-Specific;
RA MEDLINE=95215351; PubMed=7700877;
RL "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35450; AAA74614.1; -
CC InterPro: IPR000221; Protamine_P1.
CC Pfam: PF00260; protamine_P1; 1.
CC PROSITE: PS00048; Protamine_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
KW INIT_MET 0 BY SIMILARITY.
FT SEQUENCE 61 AA; 8495 MW; 58C2925C80C49A1C CRC64;
SQ
Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 RRLSYRRRRS 11
DB 44 RRGYSRRRYS 54
RESULT 9
HSP1_MACEU STANDARD; PRT; 61 AA.
ID HSP1_MACEU
AC P42141; P42153;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1
OS Macropus rufogriseus (Red-necked wallaby), and
OS Wallabia bicolor (Swamp wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9320, 9330;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE-Specific;
RA MEDLINE=95215351; PubMed=7700877;
RL "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35329; AAA74610.1; -
CC EMBL; L35328; AAA74609.1; -
CC InterPro: IPR000221; Protamine_P1.
CC Pfam: PF00260; protamine_P1; 1.
CC PROSITE: PS00048; Protamine_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
KW INIT_MET 0 BY SIMILARITY.
FT SEQUENCE 61 AA; 8571 MW; B822925C80C490BC CRC64;
SQ
Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 RRLSYRRRRS 11
DB 44 RRGYSRRRYS 54
```

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```
CC EMBL; L35450; AAA74614.1; -
CC InterPro: IPR000221; Protamine_P1.
CC Pfam: PF00260; protamine_P1; 1.
CC PROSITE: PS00048; Protamine_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
KW INIT_MET 0 BY SIMILARITY.
FT SEQUENCE 61 AA; 8495 MW; 58C2925C80C49A1C CRC64;
SQ
Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 RRLSYRRRRS 11
DB 44 RRGYSRRRYS 54
RESULT 9
HSP1_MACEU STANDARD; PRT; 61 AA.
ID HSP1_MACEU
AC P42141; P42153;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1
OS Macropus rufogriseus (Red-necked wallaby), and
OS Wallabia bicolor (Swamp wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9320, 9330;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE-Specific;
RA MEDLINE=95215351; PubMed=7700877;
RL "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35329; AAA74610.1; -
CC EMBL; L35328; AAA74609.1; -
CC InterPro: IPR000221; Protamine_P1.
CC Pfam: PF00260; protamine_P1; 1.
CC PROSITE: PS00048; Protamine_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
KW INIT_MET 0 BY SIMILARITY.
FT SEQUENCE 61 AA; 8571 MW; B822925C80C490BC CRC64;
SQ
Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 RRLSYRRRRS 11
DB 44 RRGYSRRRYS 54
```

```

Db 44 RRCYSRRRYS 54

RESULT 10
HSP1_PARB1
ID HSP1_PARB1 STANDARD; PRT; 61 AA.
AC O18768;
DT 13-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Parantechinus bilarni (Broad-footed marsupial mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Parantechinus.
OX NCBI_TaxID=32555;
RN [1]
RP SEQUENCE FROM N.A.
RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
RL cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35324; AAA74608.1; -
DR EMBL; AF010277; AAB69307.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine.P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 61 AA; 8421 MW; C02857DF087FC9A9 CRC64;

Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRFS 11
II |||||:|
Db 45 RRCYSRRRYS 55

RESULT 11
HSP1_SARHA
ID HSP1_SARHA STANDARD; PRT; 61 AA.
AC P42151;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Sarcophilus harrisii (Tasmanian devil), and
OS Dasyurus maculatus (Tiger quoll).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sarcophilus.
OX NCBI_TaxID=9305, 9281;
RN [1]
RP SEQUENCE FROM N.A.
RA SPECIES-S.harrisii; TISSUE-Sperm;
RX MEDLINE=95215351; PubMed=7700877;

Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRFS 11
II |||||:|
Db 45 RRCYSRRRYS 55

RESULT 12
HSP1_TRIVU
ID HSP1_TRIVU STANDARD; PRT; 61 AA.
AC P42152;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxID=9337;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).
RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-D.maculatus;
RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
RL cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35324; AAA74608.1; -
DR EMBL; AF010276; AAB69306.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine.P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 61 AA; 8410 MW; 4A215D3D85E71230 CRC64;

Query Match 57.7%; Score 41; DB 1; Length 61;
Best Local Similarity 72.7%; Pred. No. 0.41;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRFS 11
II |||||:|
Db 44 RRCYSRRRYS 54

RESULT 12
HSP1_TRIVU
ID HSP1_TRIVU STANDARD; PRT; 61 AA.
AC P42152;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxID=9337;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).
RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L32744; AAA99479.1; -  
InterPro; IPR000221; Protamine\_P1.  
Pfam; PF00260; protamine\_P1; 1.  
PROSITE; PS00048; NUCLEOSOME CORE; Spermatogenesis; DNA-binding;  
Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
Testis; DNA condensation; Nuclear protein.  
INIT MET 0 0 BY SIMILARITY.  
SEQUENCE 61 AA; 8571 MW; 802287E527EE816C CRC64;

Query Match 57.7%; Score 41; DB 1; Length 61;  
Best Local Similarity 72.7%; Pred. No. 0.41; Indels 2; Gaps 0;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRPS 11  
DB 44 RRGYSRRRYS 54

RESULT 13  
HSPL\_DASRO STANDARD; PRT; 62 AA.  
ID HSPL\_DASRO STANDARD; PRT; 62 AA.  
AC P42134; P42144; P42149;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Dasykaluta rosamondae, Parantechinus apicalis (Dibbler), and  
OS Pseudantechinus macdonnellensis (Fat-tailed marsupial mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasykaluta.  
OC NCBI\_TaxID=33560, 9291, 9299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Sperm;  
RC MEDLINE=95215351; PubMed=7700877;  
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;  
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L35325; AAA74605.1; -  
EMBL; L35326; AAA74607.1; -  
EMBL; L35337; AAA74603.1; -  
InterPro; IPR000221; Protamine\_P1.  
Pfam; PF00260; protamine\_P1; 1.  
PROSITE; PS00048; PROTAMINE\_P1; 1.  
Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
Testis; DNA condensation; Nuclear protein.  
INIT MET 0 0 BY SIMILARITY.  
SEQUENCE 62 AA; 8585 MW; 99C02857DF087FC9 CRC64;

Query Match 57.7%; Score 41; DB 1; Length 62;  
Best Local Similarity 72.7%; Pred. No. 0.42;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RRLSYRRRPS 11  
DB 45 RRGYSRRRYS 55

RESULT 14  
HSPL\_MURLO STANDARD; PRT; 62 AA.  
ID HSPL\_MURLO STANDARD; PRT; 62 AA.  
AC P42140; P42150; P42154;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Murexia longicaudata,  
OS Phascogale tapoatafa (Common wambenger),  
OS Smynthopsis brasiicaudata (Fat-tailed dunnart),  
OS Myrmecobius fasciatus (Numbat), and  
OS Thylacinus cynocephalus (Tasmanian wolf).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Murexia.  
OC NCBI\_TaxID=37736, 9293, 9301, 55782, 9275;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Sperm;  
RC MEDLINE=95215351; PubMed=7700877;  
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;  
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
CC [2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RX SPECIES=M. fasciatus, and T. cynocephalus;  
RC MEDLINE=97368867; PubMed=9225481;  
RA Krajewski C., Buckley L., Westernman M.;  
RT "DNA phylogeny of the marsupial wolf resolved."  
RL Proc. R. Soc. Lond., B, Biol. Sci. 264:911-917(1997).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L35336; AAA74600.1; -  
EMBL; L35337; AAA74606.1; -  
EMBL; L32743; AAA99478.1; -  
EMBL; U87139; AAB91327.1; -  
EMBL; U87140; AAB91328.1; -  
InterPro; IPR000221; Protamine\_P1.  
Pfam; PF00260; protamine\_P1; 1.  
PROSITE; PS00048; PROTAMINE\_P1; 1.  
Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
Testis; DNA condensation; Nuclear protein.  
INIT MET 0 0 BY SIMILARITY.  
SEQUENCE 62 AA; 8566 MW; 99C02857CBB73429 CRC64;

Query Match 57.7%; Score 41; DB 1; Length 62;  
Best Local Similarity 72.7%; Pred. No. 0.42; Indels 2; Gaps 0;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRPS 11  
DB 45 RRGYSRRRYS 55

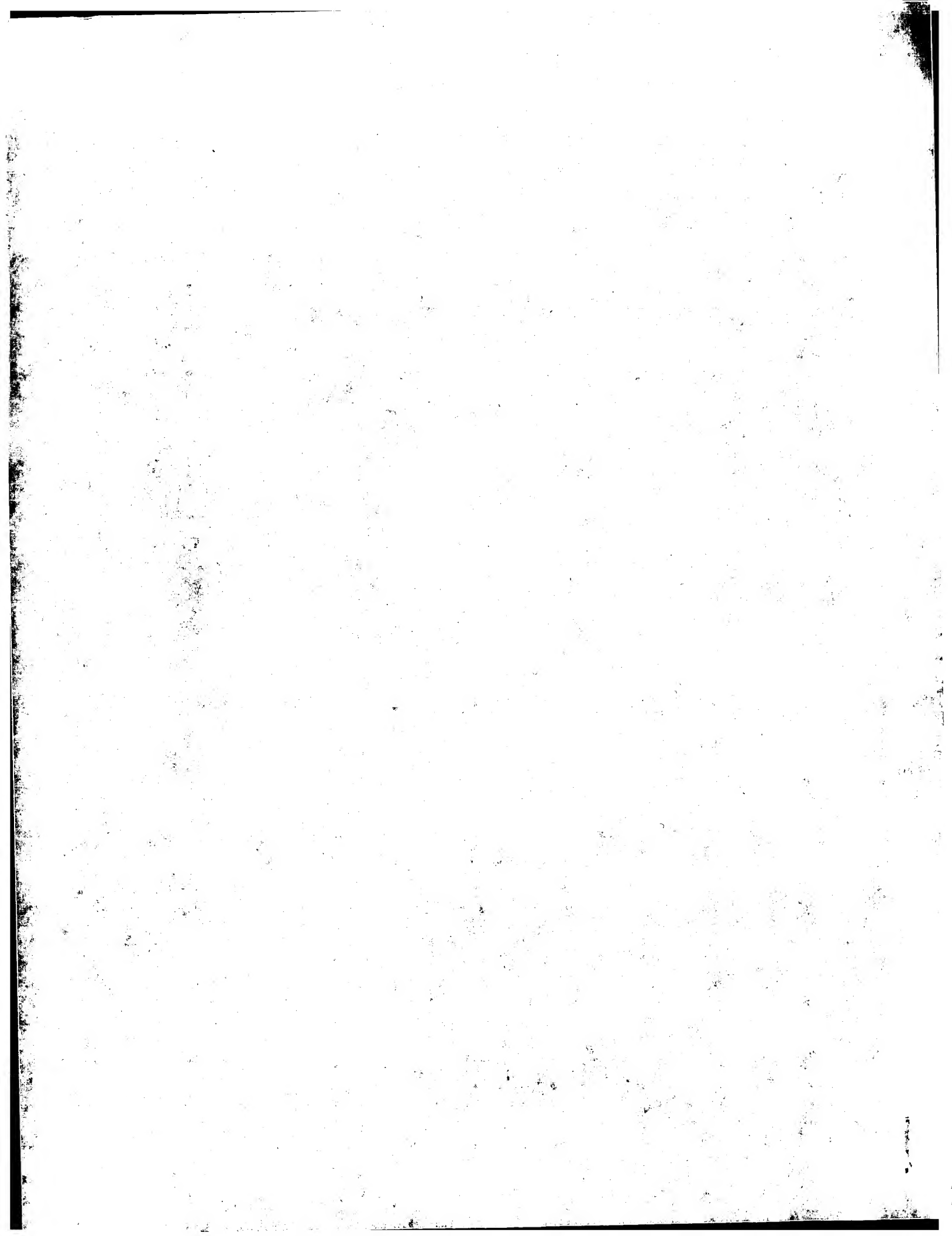


```
RESULT 15
HSPI_ANTST STANDARD; PRT; 63 AA.
AC P42129;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Antechinus stuartii (Brown marsupial mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.
OX NCBI_TaxID=9283;
RN [1]
RP SEQUENCE FROM N.A.
RC
RT MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35335; AAB95428.1;
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 63 AA; 8722 MW; D4FF92DAAB56D61 CRC64;
```

Query Match 57.7%; Score 41; DB 1; Length 63;  
Best Local Similarity 72.7%; Pred. No. 0.43;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 RRLSYRRRRFS 11  
|||  
Db 46 RRGYSRRRRYS 56

Search completed: February 12, 2002, 12:39:52  
Job time: 805 sec



RESULT	2	
Q9KK09		
ID	Q9KK09	PRELIMINARY;
AC	Q9KK09;	PRT; 580 AA.

Wed Feb 13 07:52:14 2002

us-09-485-571-24.rspt

DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE SURFACE PROTEIN PSpC.  
 GN PCPC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Iannelli F., Oggioni M.R., Spinosa M.R., Pozzi G.;  
 RT "Allelic variants of surface protein PspC (SpsA, CbpA, PbcA) of  
 RT Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF154045; AAF73826.1; -;  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SQ SEQUENCE 580 AA; 64957 MW; 272A0D8307A3E94C CRC64;

Query Match 62.0%; Score 44; DB 2; Length 580;  
 Best Local Similarity 64.3%; Pred. No. 10;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLSYRRRFSVSV 14  
 ||: || |::||| |  
 Db 8 RRMYSIRKFSVGV 21

RESULT 3

Q9KK12 PRELIMINARY; PRT; 584 AA.  
 ID Q9KK12  
 AC Q9KK12  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE SURFACE PROTEIN PSpC.  
 GN PSPC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Iannelli F., Oggioni M.R., Spinosa M.R., Pozzi G.;  
 RT "Allelic variants of surface protein PspC (SpsA, CbpA, PbcA) of  
 RT Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF154044; AAF73821.1; -;  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SQ SEQUENCE 584 AA; 65449 MW; D309081C74A24D08 CRC64;

Query Match 62.0%; Score 44; DB 2; Length 584;  
 Best Local Similarity 64.3%; Pred. No. 10;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLSYRRRFSVSV 14  
 ||: || |::||| |  
 Db 8 RRMYSIRKFSVGV 21

RESULT 4

Q9KK45

ID Q9KK45 PRELIMINARY; PRT; 587 AA.  
 AC Q9KK45;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE SURFACE PROTEIN PSpC.  
 GN PSPC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Iannelli F., Oggioni M.R., Spinosa M.R., Pozzi G.;  
 RT "Allelic variants of surface protein PspC (SpsA, CbpA, PbcA) of  
 RT Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF154014; AAF73782.1; -;  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SQ SEQUENCE 587 AA; 65798 MW; 1F7FB5ACD988299E CRC64;

Query Match 62.0%; Score 44; DB 2; Length 587;  
 Best Local Similarity 64.3%; Pred. No. 10;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLSYRRRFSVSV 14  
 ||: || |::||| |  
 Db 8 RRMYSIRKFSVGV 21

RESULT 5

Q9KK25 PRELIMINARY; PRT; 763 AA.  
 ID Q9KK25  
 AC Q9KK25;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE SURFACE PROTEIN PSpC.  
 GN PSPC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Iannelli F., Oggioni M.R., Spinosa M.R., Pozzi G.;  
 RT "Allelic variants of surface protein PspC (SpsA, CbpA, PbcA) of  
 RT Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF154033; AAF73803.1; -;  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SQ SEQUENCE 763 AA; 85198 MW; 0FE48A286F6EB197 CRC64;

Query Match 62.0%; Score 44; DB 2; Length 763;  
 Best Local Similarity 64.3%; Pred. No. 13;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLSYRRRFSVSV 14  
 ||: || |::||| |  
 Db 8 RRMYSIRKFSVGV 21

RESULT 6



us-09-485-571-24.rspt

Wed Feb 13 07:52:14 2002

```

Q9GLQ3;
AC 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Macropus parryi (Whiptail wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9318;
RN [1]
RP SEQUENCE FROM N.A.
RA Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
CC EMBL; AF187533; AAG27950.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 61 AA; 8500 MW; A07F5C81C4664B6F CRC64;

Query Match 57.7%; Score 41; DB 6; Length 61;
Best Local Similarity 72.7%; Pred. No. 3.7;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLYSRRRFS 11
DB 44 RRGYSRRRYS 54

RESULT 10
Q9GLQ3 PRELIMINARY; PRT; 61 AA.
AC Q9GLQ3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Onychogalea fraenata (bridled nail-tailed wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Onychogalea.
OX NCBI_TaxID=114227;
RN [1]
RP SEQUENCE FROM N.A.
RA Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
CC EMBL; AF187542; AAG27959.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 61 AA; 8454 MW; CBB8EFC966E44B6A CRC64;

Query Match 57.7%; Score 41; DB 6; Length 61;
Best Local Similarity 72.7%; Pred. No. 3.7;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRLYSRRRFS 11
DB 44 RRGYSRRRYS 54

RESULT 11
Q9GLP9 PRELIMINARY; PRT; 61 AA.
AC Q9GLP9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Aepyprymnus rufescens (rufous rat-kangaroo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Aepyprymnus.
OX NCBI_TaxID=38598;
RN [1]
RP SEQUENCE FROM N.A.
RA Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
CC EMBL; AF187547; AAG27964.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 61 AA; 8576 MW; 7B324691290717AD CRC64;

Query Match 57.7%; Score 41; DB 6; Length 61;
Best Local Similarity 72.7%; Pred. No. 3.7;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLYSRRRFS 11
DB 45 RRGYSRRRYS 55

RESULT 12
Q9GLP8 PRELIMINARY; PRT; 61 AA.
AC Q9GLP8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Potorous longipes (long-footed potoroo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Potorous.
OX NCBI_TaxID=55310;
RN [1]
RP SEQUENCE FROM N.A.
RA Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
CC EMBL; AF187548; AAG27965.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.

```

SO SEQUENCE 61 AA; 8431 MW; D745F1F638DBDCDC CRC64;

Query Match 57.7%; Score 41; DB 6; Length 61;  
Best Local Similarity 72.7%; Pred. No. 3.7;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRFS 11  
|| |||||:  
Db 45 RRGYSRRRYS 55

RESULT 13

ID Q9GJQ1 PRELIMINARY; PRT; 61 AA.

AC Q9GJQ1;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE SPERM PROTAMINE P1.

GN PRM1.  
OS Onychogalea unguifera (northern nail-tailed wallaby),  
OS Dendrolagus goodfellowi (Goodfellow's tree kangaroo), and  
OS Setonix brachyurus (quokka).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Onychogalea.

OX NCBI\_TaxID=65626, 69260, 30670;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=O.unguifera, D.goodfellowi, and S.brachyurus;

RA Burk A., Springer M.S.;

RT "The Chronicle of Kangaroo Evolution.";

RL J. Mammal. Evol. 0:0-0(2000).

CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY  
CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).

CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.

DR EMBL: AF187543; AAG27960.1; -

DR EMBL: AF187537; AAG27954.1; -

DR EMBL: AF187541; AAG27958.1; -

DR InterPro: IPR000221; Protamine\_P1.

DR Pfam: PF00260; protamine\_P1; 1.

DR PROSITE: PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; DNA condensation; Nuclear protein;

KW Nucleosome core; Spermatogenesis; Testis.

SO SEQUENCE 61 AA; 8546 MW; F0D55C81C4664B62 CRC64;

Query Match 57.7%; Score 41; DB 6; Length 61;  
Best Local Similarity 72.7%; Pred. No. 3.7;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRFS 11  
|| |||||:  
Db 44 RRGYSRRRYS 54

RESULT 14

ID Q9TUC2 PRELIMINARY; PRT; 62 AA.

AC Q9TUC2;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE SPERM PROTAMINE P1.

GN PRM1.

OS Sminthopsis longicaudata (long-tailed dunnart).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sminthopsis.

OX NCBI\_TaxID=90764;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99310778; PubMed=10381317;  
RA Blacket M.J., Krajewski C., Labrinidis A., Cambron B., Cooper S.,  
RA Westerman M.;  
RT "Systematic relationships within the dasyurid marsupial tribe  
RT Sminthopsini--a multigene approach.";  
RL Mol. Phylogenet. Evol. 12:140-155(1999).  
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY  
CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).

CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.

DR EMBL: AF089881; AAD5340.1; -

DR InterPro: IPR000221; Protamine\_P1.

DR Pfam: PF00260; protamine\_P1; 1.

DR PROSITE: PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; DNA condensation; Nuclear protein;

KW Nucleosome core; Spermatogenesis; Testis.

SO SEQUENCE 62 AA; 8640 MW; A9EEA7D7C77964A9 CRC64;

Query Match 57.7%; Score 41; DB 6; Length 62;  
Best Local Similarity 72.7%; Pred. No. 3.8;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRFS 11  
|| |||||:  
Db 46 RRGYSRRRYS 56

RESULT 15

ID Q9GLQ7 PRELIMINARY; PRT; 62 AA.

AC Q9GLQ7;

DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE PROTAMINE P1.

OS Petrogale xanthopus (Ring-tailed rock wallaby).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Petrogale.

OX NCBI\_TaxID=9325;

RN [1]

RP SEQUENCE FROM N.A.

RA Burk A., Springer M.S.;

RT "The Chronicle of Kangaroo Evolution.";

RL J. Mammal. Evol. 0:0-0(2000).

DR EMBL: AF187535; AAG27952.1; -

DR InterPro: IPR000221; Protamine\_P1.

DR Pfam: PF00260; protamine\_P1; 1.

DR PROSITE: PS00048; PROTAMINE\_P1; UNKNOWN.1.

SO SEQUENCE 62 AA; 8656 MW; 8EBE685C8089D007 CRC64;

Query Match 57.7%; Score 41; DB 6; Length 62;  
Best Local Similarity 72.7%; Pred. No. 3.8;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRFS 11  
|| |||||:  
Db 45 RRGYSRRRYS 55

Search completed: February 12, 2002, 12:38:40  
Job time: 753 sec

us-09-485-571-24.rspt

Wed Feb 13 07:52:14 2002

-----



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:32 ; Search time 242.57 Seconds  
(without alignments)  
5.497 Million cell updates/sec

Title: US-09-485-571-25

Perfect score: 91

Sequence: 1 RGRRLSYRRRSTSTGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_ll01.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	91	100.0	18	AAW99412	Protegrin derivati
2	91	100.0	18	AAV93177	Protegrin-like pep
3	91	100.0	18	AAV93179	Protegrin-like pep
4	91	100.0	18	AAV93615	Peptide which may
5	81	89.0	18	AAW99403	Protegrin derivati
6	81	89.0	18	AAV93616	Peptide which may
7	67	73.6	18	AAW18153	Cationic, antimicr
8	66	72.5	18	AAV22005	Antimicrobial pept
9	66	72.5	18	AAV22007	Antimicrobial pept
10	65	71.4	18	AAW36429	Antimicrobial prot
11	65	71.4	18	AAW09084	Cationic, antimicr

12	65	71.4	18	AAW09085	Cationic, antimicr
13	64	70.3	18	AAW18151	Cationic, antimicr
14	64	70.3	18	AAW18152	Cationic, antimicr
15	63	69.2	18	AAW18150	Cationic, antimicr
16	63	69.2	18	AAW18151	Protegrin PG-1. S
17	61	67.0	18	AAW78751	Protegrin peptide
18	61	67.0	18	AAW78776	Antimicrobial prot
19	61	67.0	18	AAW36208	Antimicrobial prot
20	61	67.0	18	AAW36322	Antimicrobial prot
21	61	67.0	18	AAW36353	Antimicrobial prot
22	61	67.0	18	AAW35578	Cationic, antimicr
23	61	67.0	18	AAW18144	Cationic, antimicr
24	61	67.0	18	AAW18130	Cationic, antimicr
25	61	67.0	18	AAW09073	Porcine protegrin
26	61	67.0	18	AAW66458	Cationic peptide p
27	61	67.0	18	AAV22018	Protegrin peptide
28	61	67.0	18	AAV93170	Protegrin peptide
29	61	67.0	18	AAV93608	Protegrin peptide
30	61	67.0	18	AAV81680	Protegrin peptide
31	61	67.0	18	AAV91757	Cationic peptide P
32	61	67.0	18	AAV91843	Antimicrobial pept
33	61	67.0	18	AAV35050	Porcine protegrin-
34	61	67.0	149	AAW25081	Antimicrobial comp
35	61	67.0	149	AAW09087	Antimicrobial prot
36	60	65.9	18	AAW78773	Protegrin peptide
37	60	65.9	18	AAW78765	Protegrin peptide
38	60	65.9	18	AAW18147	Cationic, antimicr
39	60	65.9	18	AAW18148	Cationic, antimicr
40	60	65.9	18	AAW18149	Cationic, antimicr
41	59	64.8	18	AAW23754	Tachytegrin and/or
42	59	64.8	18	AAW23732	Tachytegrin and/or
43	59	64.8	18	AAW23762	Tachytegrin and/or
44	59	64.8	18	AAW18128	Cationic, antimicr
45	59	64.8	18	AAW18129	Cationic, antimicr

ALIGNMENTS

RESULT 1

AAW99412 standard; peptide; 18 AA.

ID XX

AC AAW99412;

DT 08-JUN-1999 (first entry)

XX Protegrin derivative peptide SM2196.

DE Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;

XX disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;

KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;

KW nucleus; blood-brain barrier.

XX Synthetic.

OS WO9907728-A2.

PN 18-FEB-1999.

XX 06-AUG-1998; 98WO-FR01757.

XX 12-AUG-1997; 97FR-0010297.

XX (SYNT-) SYNT:EM SA.

XX Calas E, Chavanieu A, Grassy G, Kaczorek M;

XX WPI; 1999-190034/16.

XX Derivatives of antibiotic peptides lacking disulfide bridges - used

PT as carriers to deliver active agents into cells

XX

wed Feb 13 07:52:15 2002

Claim 7; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the protegrin family of  
XX peptide antibiotics. Protegrin antibiotics form part of the peptide  
XX antibiotic family which contain a beta-sheet secondary structure linked  
XX by disulphide bridges. The new derivatives are linear and lack the  
XX disulphide bridge. The novel derivatives are used to deliver active  
XX agents to an organism, e.g. therapeutic proteins, antibodies (or their  
XX fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
XX antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
XX and non-lytic but can cross mammalian cell membranes rapidly by a passive  
XX mechanism, so can deliver active agents to cytoplasm and nucleus,  
XX including crossing the blood-brain barrier.

SX Sequence 18 AA;

Query Match 100.0%; Score 91; DB 20; Length 18;  
Best Local Similarity 100.0%; Pred. NO. 3.5e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRRLSYRRRSTSTGR 18  
Db 1 RGRRLSYRRRSTSTGR 18

RESULT 2  
AA93177 2  
ID AA93177 standard; peptide; 18 AA.

AC AA93177;

DT 06-DEC-2000 (first entry)

XX Protegrin-like peptide antibiotic Doxo-SynBI.

DE Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
XX blood-brain barrier; diagnostic; central nervous system; protegrin;  
KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
KW cancer; Parkinson's disease; depression; pain; meningitis.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "linked to doxorubicin via a succinate  
FT (-CO-(CH<sub>2</sub>)<sub>2</sub>-CO-) linker; optionally linked  
FT to benzylpenicillin by a glycoamide linker"

XX WO200032236-A1.

XX 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02938.

XX 30-NOV-1998; 98FR-0015074.

XX (SYNT-) SYNT:EM SA.

XX Clair P, Kaczorek M, Tamsamani J;

XX WPI; 2000-422871/36.

XX Use of linear peptides as vectors for active ingredients, useful for  
XX diagnosis and treatment of central nervous system diseases, can  
XX transport agents passively across the blood-brain barrier

XX Example I; Page 13; 54pp; French.

XX The invention relates to the use of linear peptides, coupled to an active  
XX agent, to prepare a composition able to cross the blood-brain barrier  
XX for diagnosis or treatment of disorders localised in the central nervous  
XX system. The linear peptide preferably has the formula: (a) XI-X16;

CC (b); BXXBXXXXBXXXXXB; or (c) BXXBXXXXBXXXXBXXB, where: each of X1-X16  
CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
CC be Trp; each B is aa containing a side chain that includes a basic group;  
CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
CC Peptides able to cross the BBB include protegrins, Antennapedia  
CC tachyplesins, transportin, etc. Of these several families have cytolytic  
CC effects and are termed peptide antibiotics. They fall into 3 main  
CC categories based on their structure: (i) peptides with alpha-helices,  
CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
CC beta-sheets, e.g. protegrin, tachyplesins, defensins; (iii) peptides  
CC with no major structure but containing bends due to the presence of  
CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
CC fall into the peptide antibiotic categories defined above: (a)-peptides  
CC are based on the Antennapedia family peptides; (b)-peptides are based on  
CC protegrins; and (c)-peptides are based on tachyplesins. This sequence  
CC represents a synthetic linear peptide designed on peptides able to cross  
CC the BBB and is conjugated to a doxorubicin molecule by a succinate  
CC linker. The peptide may also be linked to a benzylpenicillin molecule  
CC by a glycoamide linker.  
CC Conjugates of the linear peptides and the active agent are particularly  
CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
CC Parkinson's diseases, depression, pain and meningitis, but also for  
CC studying drug behaviour in BBB models.

XX Sequence 18 AA;

Query Match 100.0%; Score 91; DB 21; Length 18;  
Best Local Similarity 100.0%; Pred. NO. 3.5e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRRLSYRRRSTSTGR 18  
Db 1 RGRRLSYRRRSTSTGR 18

RESULT 3  
AA93179 3  
ID AA93179 standard; peptide; 18 AA.

AC AA93179;

XX 06-DEC-2000 (first entry)

XX Protegrin-like peptide antibiotic Dal-SynBI.

XX Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
KW cancer; Parkinson's disease; depression; pain; meningitis; dalargin.

XX Synthetic.

XX Key Location/Qualifiers

FT Cross-links 1 /note= "cross-links to a molecule of dalargin via  
FT a disulphide linker"

XX WO200032236-A1.

XX 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02938.

XX 30-NOV-1998; 98FR-0015074.

XX (SYNT-) SYNT:EM SA.

XX Clair P, Kaczorek M, Tamsamani J;

XX WPI; 2000-422871/36.



CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory agents, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.  
 XX  
 XX Sequence 18 AA;

Query Match 89.0%; Score 81; DB 20; Length 18;  
 Best Local Similarity 88.9%; Pred. No. 1.5e-06;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RGRRLSYRRRSTSTGR 18  
 |||||  
 DB 1 rgrrlsyrrrfsvsgr 18

RESULT 6  
 AAY93616  
 ID AAY93616 standard; peptide; 18 AA.  
 XX  
 AC AAY93616;  
 XX  
 DT 25-SEP-2000 (first entry)

XX Peptide which may be linked to anticancer agents.  
 DE  
 XX Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 KW cancer.  
 XX  
 OS Unidentified.  
 XX  
 XX WO200032237-A1.  
 PN  
 XX 08-JUN-2000.  
 PD  
 XX 26-NOV-1999; 99WO-FR02939.  
 PF  
 XX 30-NOV-1998; 98FR-0015073.  
 PR  
 XX (SYNT-) SYNT:EM SA.  
 PA  
 XX Tensamani J, Kaczorek M, Colin De Verdiere A;  
 PI  
 XX WPI; 2000-412166/35.  
 DR  
 XX New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells  
 XX  
 PS Disclosure; Page 8; 34pp; French.

XX The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.  
 XX  
 XX Sequence 18 AA;

Query Match 89.0%; Score 81; DB 21; Length 18;  
 Best Local Similarity 88.9%; Pred. No. 1.5e-06;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 RGRRLSYRRRSTSTGR 18  
 |||||  
 DB 1 rgrrlsyrrrfsvsgr 18

RESULT 7  
 AAW18153  
 ID AAW18153 standard; peptide; 18 AA.  
 XX  
 AC AAW18153;  
 XX  
 DT 11-AUG-1997 (first entry)

XX Cationic, antimicrobial, virus-neutralising proteogrin PC-57.  
 DE  
 XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW Candida albicans; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
 KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;  
 KW food.  
 XX  
 OS Synthetic.  
 XX  
 XX WO9637508-A1.  
 PN  
 XX 28-NOV-1996.  
 PD  
 XX 24-MAY-1996; 96WO-US07594.  
 PF  
 XX 07-JUL-1995; 95US-0499523.  
 PR  
 XX 26-MAY-1995; 95US-0451832.  
 PR  
 XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.  
 PA  
 XX Harwig SSL, Kokryakov VN, Lehrer RI;  
 PI  
 XX WPI; 1997-033984/03.

XX Cationic, antimicrobial, virus-neutralising proteogrin peptide(s) -  
 PT useful for the treatment of microbial infection, as food  
 PT preservatives and in eye care solutions  
 XX  
 XX Claim 6; Page 64; 106pp; English.

XX The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC cysteine bridge. This peptide is in a form where all the  
 CC cysteine residues are replaced by a hydrophobic, small or large polar  
 CC amino acid (e.g. alanine in this case). Peptides of this formula are  
 CC designated proteogrin and are useful as anti-bacterial, anti-viral and  
 CC anti-fungal agents in plants and animals. The proteogrin confer  
 CC resistance to microbial or viral infection in plants by preventing the  
 CC growth of a virus or microbe and inactivate the endotoxin of gram-  
 CC negative bacteria. The proteogrin are particularly useful for the  
 CC treatment of sexually transmitted disease caused by microorganisms e.g.  
 CC Candida albicans, HIV-1, Chlamydia trachomatis, Treponema pallidum and  
 CC Neisseria gonorrhoeae. They can also be used in eye care solutions and  
 CC as preservatives for food. The proteogrin are more effective under  
 CC physiological conditions (e.g. in the presence of serum) than certain  
 CC antibiotics and are non-toxic to the cells of higher organisms.

Sequence 18 AA;

Query Match 73.6%; Score 67; DB 18; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.00026;  
Matches 12; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 RGRRLSYRRRFFSTGR 18  
|||||:|||||: ||  
Db 1 rgrrlayrrrrfawgr 18

RESULT 8

AAV22005  
ID AAV22005 standard; peptide: 18 AA.

AC AAV22005;

DT 23-AUG-1999 (first entry)

DE Antimicrobial peptide derivative of protegrin.

XX Antimicrobial peptide; protegrin; microbe growth inhibitor; Pseudomonas;  
KW microbial infection; Staphylococcus aureus; Helicobacter pylori; therapy;  
KW antibiotic-resistant bacterium; disinfectant composition; preservative;  
KW haemolytic activity; systemic antibiotic.

XX Sus scrofa.

XX WO9927945-A1.

XX 10-JUN-1999.

XX 01-DEC-1998; 98WO-US25458.

XX 03-DEC-1997; 97US-0984294.

XX (INTR-) INTRABIOTICS PHARM INC.

XX Chang CC, Chen J, Lehrer RI, Radel PA;

XX WPI; 1999-385321/32.

XX Antimicrobial threonine-containing protegrins

XX Claim 11; Page 52; 75pp; English.

XX This sequence represents an example of an antimicrobial peptide (I) of  
CC the invention, and comprises 10-30 amino acid residues based on a  
CC naturally-occurring protegrin peptide. The antimicrobial peptide is  
CC useful in environmental composition for application to plants or plant  
CC environments, for inhibiting the growth of microbes. The peptide is  
CC useful for treating or preventing microbial infections, e.g. caused by  
CC Staphylococcus aureus, Pseudomonas, Helicobacter pylori or an  
CC antibiotic-resistant bacterium, or related diseases. The peptide can be  
CC used to inactivate the endotoxin of Gram-negative bacteria. The peptides  
CC can be used in disinfectant compositions, and as preservatives for  
CC materials such as foodstuffs, cosmetics, medicaments or other materials  
CC containing nutrients for organisms. The peptides exhibit decreased  
CC haemolytic activity against human red blood cells as compared with native  
CC PG-1 and melittin. They have improved serum compatibility and therefore  
CC improved use as systemic antibiotics. At the same time the peptides  
CC provide broad spectrum activity with a low frequency of resistance.

XX Sequence 18 AA;

Query Match 72.5%; Score 66; DB 20; Length 18;  
Best Local Similarity 72.2%; Pred. No. 0.00038;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 RGRRLSYRRRFFSTGR 18  
|||||:|||||: ||  
Db 1 rgrrlcyrrrrfctcgr 18

RESULT 9

AAV22007  
ID AAV22007 standard; peptide: 18 AA.

AC AAV22007;

DT 23-AUG-1999 (first entry)

DE Antimicrobial peptide derivative of protegrin.

XX Antimicrobial peptide; protegrin; microbe growth inhibitor; Pseudomonas;  
KW microbial infection; Staphylococcus aureus; Helicobacter pylori; therapy;  
KW antibiotic-resistant bacterium; disinfectant composition; preservative;  
KW haemolytic activity; systemic antibiotic.

XX Sus scrofa.

XX WO9927945-A1.

XX 10-JUN-1999.

XX 01-DEC-1998; 98WO-US25458.

XX 03-DEC-1997; 97US-0984294.

XX (INTR-) INTRABIOTICS PHARM INC.

XX Chang CC, Chen J, Lehrer RI, Radel PA;

XX WPI; 1999-385321/32.

XX Antimicrobial threonine-containing protegrins

XX Claim 11; Page 52; 75pp; English.

XX This sequence represents an example of an antimicrobial peptide (I) of  
CC the invention, and comprises 10-30 amino acid residues based on a  
CC naturally-occurring protegrin peptide. The antimicrobial peptide is  
CC useful in environmental composition for application to plants or plant  
CC environments, for inhibiting the growth of microbes. The peptide is  
CC useful for treating or preventing microbial infections, e.g. caused by  
CC Staphylococcus aureus, Pseudomonas, Helicobacter pylori or an  
CC antibiotic-resistant bacterium, or related diseases. The peptide can be  
CC used to inactivate the endotoxin of Gram-negative bacteria. The peptides  
CC can be used in disinfectant compositions, and as preservatives for  
CC materials such as foodstuffs, cosmetics, medicaments or other materials  
CC containing nutrients for organisms. The peptides exhibit decreased  
CC haemolytic activity against human red blood cells as compared with native  
CC PG-1 and melittin. They have improved serum compatibility and therefore  
CC improved use as systemic antibiotics. At the same time the peptides  
CC provide broad spectrum activity with a low frequency of resistance.

XX Sequence 18 AA;

Query Match 72.5%; Score 66; DB 20; Length 18;  
Best Local Similarity 72.2%; Pred. No. 0.00038;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 RGRRLSYRRRFFSTGR 18  
|||||:|||||: ||  
Db 1 rgrrlcyrrrrfctcgr 18

RESULT 10

AAW36429  
ID AAW36429 standard; peptide: 18 AA.

AC AAW36429;

DT 13-FEB-1998 (first entry)

Wed Feb 13 07:52:15 2002

XX DE Antimicrobial protegrin peptide (229).

XX Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;

KW gram-negative; bacteria; yeast; fungus; protozoa; virus;

KW retrovirus; HIV; human immunodeficiency virus; preservation;

KW disinfection; prophylaxis; treatment; infection; disease;

KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;

KW Helicobacter pylori; sexually transmitted disease; oral mucositis;

KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;

KW respiratory infection; urinary tract infection; MRSA; protozoan;

KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;

KW penicillin resistant Streptococcus pneumoniae; pig; porcine;

KW methicillin resistant Staphylococcus aureus; systemic candidiasis.

XX Synthetic.

OS Sus scrofa.

XX WO9718826-A1.

XX 29-MAY-1997.

XX 22-NOV-1996; 96WO-US18544.

XX 21-NOV-1996; 96US-0752852.

PR 22-NOV-1995; 95US-0562346.

PR 17-MAY-1996; 96US-0649811.

PR 01-AUG-1996; 96US-0690921.

XX (INTR-) INTRABIOTICS PHARM INC.

PA (REGC ) UNIV CALIFORNIA.

XX Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;

PI WPI; 1997-297871/27.

XX New antimicrobial protegrin peptide(s) - having activity against

PT bacteria, yeast, fungi, protozoa and certain strains of viruses

PT (e.g. HIV)

XX Claim 23; Page 110; 130pp; English.

XX The present sequence is an antimicrobial protegrin peptide, which

CC has a broad spectrum of activity against microbial targets, fungi,

CC including gram-positive and gram-negative bacteria, yeast, fungi,

CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.

CC It can be used to preserve or disinfect a variety of materials,

CC including medical equipment, foodstuffs, cosmetics, contact lens

CC solutions, medicaments or other nutrient containing materials. It

CC can also be used for the prophylaxis or treatment of microbial

CC infections or diseases in plants and animals, e.g. conjunctivitis,

CC keratitis, corneal ulcers, stomach ulcers associated with

CC Helicobacter pylori, sexually transmitted diseases, gram-negative

CC sepsis, endocarditis, pneumonia and other respiratory infections,

CC urinary tract infections, systemic candidiasis and oral mucositis.

CC It is biostatic or biocidal against clinically relevant pathogens

CC exhibiting multi-drug resistance, e.g. vancomycin resistant

CC Enterococcus faecium or faecalis, penicillin resistant

CC Streptococcus pneumoniae and methicillin resistant Staphylococcus

CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5

CC to 1 mg/kg/day, by injection.

XX Sequence 18 AA;

XX Query Match 71.4%; Score 65; DB 18; Length 18;

Best Local Similarity 66.7%; Pred. No. 0.00056;

Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 RGGRLSYRRRFTSTGR 18

Db 1 rggrlcyarrrfvcvgr 18

RESULT 11

AAW09084

ID AAW09084 standard; peptide: 18 AA.

XX

AC AAW09084;

XX

DT 11-AUG-1997 (first entry)

DE Cationic, antimicrobial, virus-neutralising protegrin IB-288.

XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;

KW Candida albicans; gram-negative bacteria; STD;

KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;

KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;

KW food.

XX Synthetic.

XX

XX Key Location/Qualifiers

FT Modified-site 1

FT /note= "Acylation"

FT Disulfide-bond 6..15

FT /note= "Results in bullet form peptide"

FT

XX WO9637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

XX 07-JUL-1995; 95US-0499523.

PR 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX Harwig SSL, Kokryakov VN, Lehrer RI;

PI WPI; 1997-033984/03.

XX

DR Cationic, antimicrobial, virus-neutralising protegrin peptide(s)

XX useful for the treatment of microbial infection, as food

PT preservatives and in eye care solutions

PT

PT Claim 6; Page 65; 106pp; English.

XX The present sequence is a specifically claimed example of a peptide,

CC recombinantly produced, corresponding to the generic formula:

CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A16-(A17-A18)

CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a

CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;

CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino

CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino

CC acid or proline; A17 may be absent or a basic, neutral/polar,

CC hydrophobic or small amino acid; and A18 may be absent or a basic,

CC neutral/polar, hydrophobic or small amino acid. This has a charge of at

CC least +3 and its N-terminal acylated and/or C-terminal amidated or

CC esterified forms, all of which may contain a disulphide bond to give a

CC cysteine bridge. Peptides of this formula are designated protegrins and

CC are useful as anti-bacterial, anti-viral and anti-fungal agents in

CC plants and animals. The protegrins confer resistance to microbial or

CC viral infection in plants by preventing the growth of a virus or microbe

CC and inactivate the endotoxin of gram-negative bacteria. The protegrins

CC are particularly useful for the treatment of sexually transmitted

CC disease caused by microorganisms e.g. Candida albicans, HIV-1, Chlamydia

CC trachomatis, Treponema pallidum and Neisseria gonorrhoeae. They can also

CC be used in eye care solutions and as preservatives for food. The

CC protegrins are more effective under physiological conditions (e.g. in

CC the presence of serum) than certain antibiotics and are non-toxic to the

CC cells of higher organisms.

XX Sequence 18 AA;

Query Match 71.4%; Score 65; DB 18; Length 18;  
 Best Local Similarity 66.7%; Pred. No. 0.00056;  
 Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLSYSRFFSTSTGR 18  
 |||||:||||: ||  
 Db 1 rggrycyarrfvcvgr 18

RESULT 12  
 AAW09085  
 ID AAW09085 standard; peptide; 18 AA.  
 AC AAW09085;  
 XX  
 DT 11-AUG-1997 (first entry)  
 XX  
 DE Cationic, antimicrobial, virus-neutralising protegrin IB-289.  
 XX  
 KW Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW Candida albicans; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
 KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;  
 KW food.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1  
 FT Disulfide-bond /note= "Acylated"  
 FT 6..15  
 FT /note= "results in bullet form peptide"  
 FT 18  
 FT Modified-site /note= "Amidated"  
 FT  
 XX  
 PN W09637508-A1.  
 XX  
 XX 28-NOV-1996.  
 XX  
 XX 24-MAY-1996; 96WO-US07594.  
 XX  
 PR 07-JUL-1995; 95US-0499523.  
 PR 26-MAY-1995; 95US-0451832.  
 XX  
 XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.  
 XX  
 PI Harwig SSL, Kokryakov VN, Lehrer RI;  
 XX  
 XX WPI; 1997-033984/03.  
 XX  
 DR Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
 XX useful for the treatment of microbial infection, as food  
 XX preservatives and in eye care solutions  
 XX  
 PS Claim 6; Page 65; 106pp; English.  
 XX  
 CC The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. Peptides of this formula are designated protegrins and  
 CC are useful as anti-bacterial, anti-viral and anti-fungal agents in  
 CC plants and animals. The protegrins confer resistance to microbial or  
 CC viral infection in plants by preventing the growth of a virus or microbe

CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
 CC are particularly useful for the treatment of sexually transmitted  
 CC disease caused by microorganisms e.g. Candida albicans, HIV-1, Chlamydia  
 CC trachomatis, Treponema pallidum and Neisseria gonorrhoeae. They can also  
 CC be used in eye care solutions and as preservatives for food. The  
 CC protegrins are more effective under physiological conditions (e.g. in  
 CC the presence of serum) than certain antibiotics and are non-toxic to the  
 CC cells of higher organisms.  
 XX  
 SQ Sequence 18 AA;

Query Match 71.4%; Score 65; DB 18; Length 18;  
 Best Local Similarity 66.7%; Pred. No. 0.00056;  
 Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLSYSRFFSTSTGR 18  
 |||||:||||: ||  
 Db 1 rggrycyarrfvcvgr 18

RESULT 13  
 AAW18151  
 ID AAW18151 standard; peptide; 18 AA.  
 XX  
 AC AAW18151;  
 XX  
 DT 11-AUG-1997 (first entry)  
 XX  
 DE Cationic, antimicrobial, virus-neutralising protegrin PC-55.  
 XX  
 KW Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW Candida albicans; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
 KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;  
 KW food.  
 XX  
 OS Synthetic.  
 XX  
 PN W09637508-A1.  
 XX  
 XX 28-NOV-1996.  
 XX  
 XX 24-MAY-1996; 96WO-US07594.  
 XX  
 PR 07-JUL-1995; 95US-0499523.  
 PR 26-MAY-1995; 95US-0451832.  
 XX  
 XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.  
 XX  
 PI Harwig SSL, Kokryakov VN, Lehrer RI;  
 XX  
 XX WPI; 1997-033984/03.  
 XX  
 DR Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
 XX useful for the treatment of microbial infection, as food  
 XX preservatives and in eye care solutions  
 XX  
 PS Claim 6; Page 64; 106pp; English.  
 XX  
 CC The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. This peptide is in snake form where all the

Wed Feb 13 07:52:15 2002

cysteine residues are replaced by a hydrophobic, small or large polar amino acid (e.g. alanine in this case). Peptides of this formula are designated protegrins and are useful as anti-bacterial, anti-viral and anti-fungal agents in plants and animals. The protegrins confer resistance to microbial or viral infection in plants by preventing the growth of a virus or microbe and inactivate the endotoxin of gram-negative bacteria. The protegrins are particularly useful for the treatment of sexually transmitted disease caused by microorganisms e.g. *Candida albicans*, HIV-1, *Chlamydia trachomatis*, *Treponema pallidum* and *Neisseria gonorrhoeae*. They can also be used in eye care solutions and as preservatives for food. The protegrins are more effective under physiological conditions (e.g. in the presence of serum) than certain antibiotics and are non-toxic to the cells of higher organisms.

XX Sequence 18 AA;

Query Match 70.3%; Score 64; DB 18; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.00081;  
Matches 11; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRSTSTGR 18  
Db 1 Rggrlawrrrfavavgr 18

RESULT 14  
AAW18152  
ID AAW18152 standard; peptide; 18 AA.  
XX AAW18152;  
AC AAW18152;  
XX 11-AUG-1997 (first entry)  
XX Cationic, antimicrobial, virus-neutralising protegrin PC-56.  
XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
KW *Candida albicans*; gram-negative bacteria; STD;  
KW sexually transmitted disease; HIV-1; *Chlamydia trachomatis*;  
KW *Treponema pallidum*; *Neisseria gonorrhoeae*; eye care; preservative;  
KW food.  
XX Synthetic.  
XX WO9637508-A1.  
XX 28-NOV-1996.  
XX 24-MAY-1996; 96WO-US07594.  
XX 07-JUL-1995; 95US-0499523.  
XX 26-MAY-1995; 95US-0451832.  
XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.  
XX Harwig SSL, Kokryakov VN, Lehrer RI;  
XX WPI; 1997-033984/03.  
XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
PT useful for the treatment of microbial infection, as food  
PT preservatives and in eye care solutions  
XX Claim 6; Page 64; 106pp; English.  
XX The present sequence is a specifically claimed example of a peptide,  
CC recombinantly produced, corresponding to the generic formula:  
CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
CC acid or proline; A17 may be absent or a basic, neutral/polar,

hydrophobic or small amino acid; and A18 may be absent or a basic, neutral/polar, hydrophobic or small amino acid. This has a charge of at least +3 and its N-terminal acylated and/or C-terminal amidated or esterified forms, all of which may contain a disulphide bond to give a cysteine bridge. This peptide is in snake form where all the cysteine residues are replaced by a hydrophobic, small or large polar amino acid (e.g. alanine in this case). Peptides of this formula are designated protegrins and are useful as anti-bacterial, anti-viral and anti-fungal agents in plants and animals. The protegrins confer resistance to microbial or viral infection in plants by preventing the growth of a virus or microbe and inactivate the endotoxin of gram-negative bacteria. The protegrins are particularly useful for the treatment of sexually transmitted disease caused by microorganisms e.g. *Candida albicans*, HIV-1, *Chlamydia trachomatis*, *Treponema pallidum* and *Neisseria gonorrhoeae*. They can also be used in eye care solutions and as preservatives for food. The protegrins are more effective under physiological conditions (e.g. in the presence of serum) than certain antibiotics and are non-toxic to the cells of higher organisms.

XX Sequence 18 AA;

Query Match 70.3%; Score 64; DB 18; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.00081;  
Matches 11; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRSTSTGR 18  
Db 1 Rggrlayarrwvavgr 18

RESULT 15  
AAW18150  
ID AAW18150 standard; peptide; 18 AA.  
XX AAW18150;  
AC AAW18150;  
XX 11-AUG-1997 (first entry)  
XX Cationic, antimicrobial, virus-neutralising protegrin PC-53.  
XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
KW *Candida albicans*; gram-negative bacteria; STD;  
KW sexually transmitted disease; HIV-1; *Chlamydia trachomatis*;  
KW *Treponema pallidum*; *Neisseria gonorrhoeae*; eye care; preservative;  
KW food.  
XX Synthetic.  
XX Key Location/Qualifiers  
FH Modified-site 1 /note= "Acylated"  
FT Disulfide-bond 8..13 /note= "results in kite form peptide"  
FT Modified-site 18 /note= "Amidated"  
FT WO9637508-A1.  
XX 28-NOV-1996.  
XX 24-MAY-1996; 96WO-US07594.  
XX 07-JUL-1995; 95US-0499523.  
XX 26-MAY-1995; 95US-0451832.  
XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.  
XX Harwig SSL, Kokryakov VN, Lehrer RI;  
XX WPI; 1997-033984/03.  
XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
PT



PT useful for the treatment of microbial infection, as food  
XX preservatives and in eye care solutions

PS Claim 6; Page 64; 106pp; English.

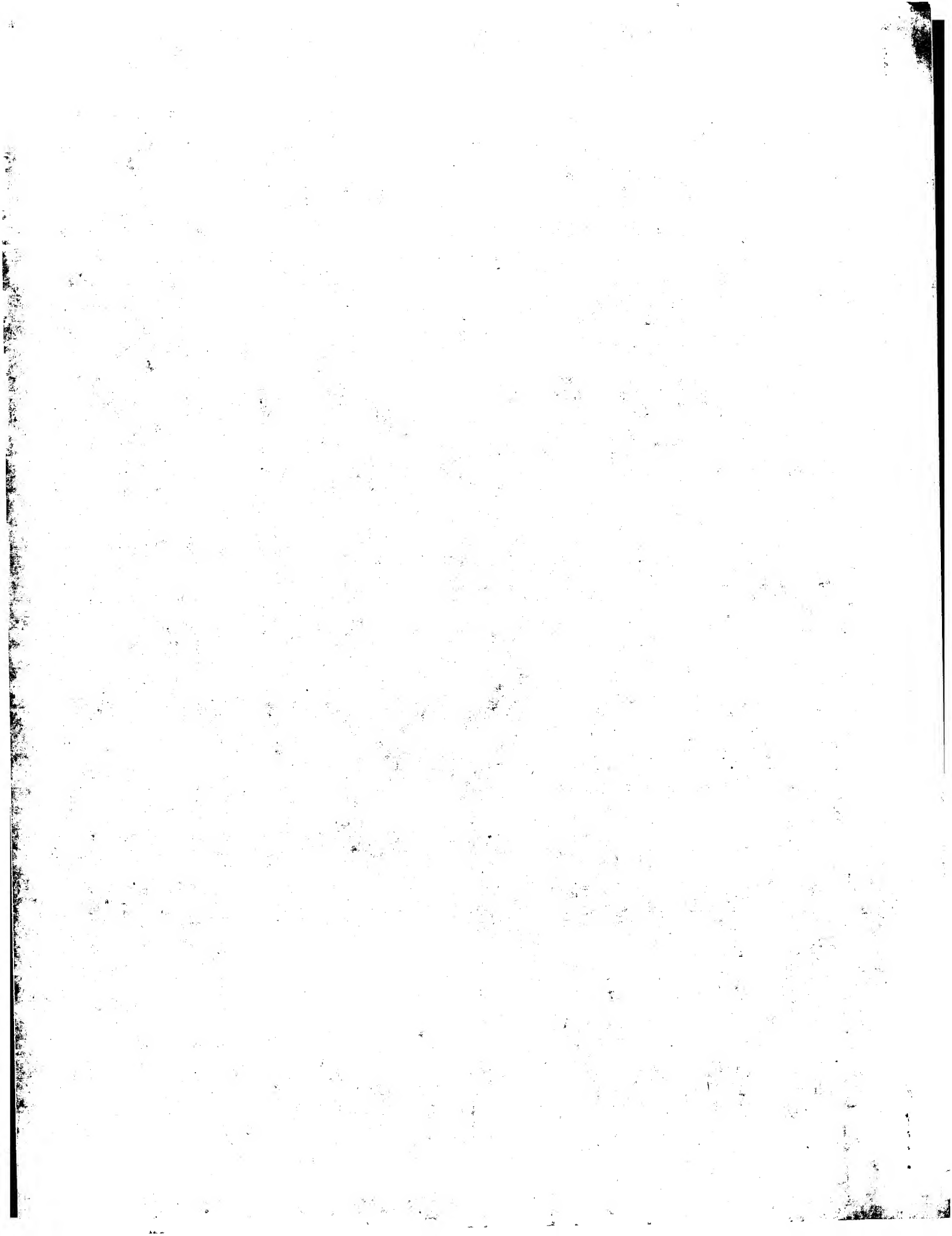
XX  
CC The present sequence is a specifically claimed example of a peptide,  
CC recombinantly produced, corresponding to the generic formula:  
CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
CC acid or proline; A17 may be absent or a basic, neutral/polar,  
CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
CC esterified forms, all of which may contain a disulphide bond to give a  
CC cysteine bridge. Peptides of this formula are designated protegrins and  
CC are useful as anti-bacterial, anti-viral and anti-fungal agents in  
CC plants and animals. The protegrins confer resistance to microbial or  
CC viral infection in plants by preventing the growth of a virus or microbe  
CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
CC are particularly useful for the treatment of sexually transmitted  
CC disease caused by microorganisms e.g. *Candida albicans*, HIV-1, *Chlamydia*  
CC *trachomatis*, *Treponema pallidum* and *Neisseria gonorrhoeae*. They can also  
CC be used in eye care solutions and as preservatives for food. The  
CC protegrins are more effective under physiological conditions (e.g. in  
CC the presence of serum) than certain antibiotics and are non-toxic to the  
CC cells of higher organisms.

XX SQ Sequence 18 AA;

Query Match 69.2%; Score 63; DB 18; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.0012;  
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 RGRRLSYRRRFTSTGR 18  
      |||||:||||:||  
Db 1 rggrlaycrrrfcvaagr 18

Search completed: February 12, 2002, 12:30:32  
Job time: 365 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:23 ; Search time 106.12 Seconds.  
(without alignments)  
3.817 Million cell updates/sec

Title: US-09-485-571-25

Perfect score: 91

Sequence: 1 RGRRLSYRRRFTSTGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	72.5	18	3	US-08-984-294-6
2	66	72.5	18	3	US-08-984-294-8
3	65	71.4	18	1	US-08-499-523-63
4	65	71.4	18	1	US-08-499-523-67
5	65	71.4	18	2	US-08-752-852A-230
6	65	71.4	18	4	US-09-128-345-63
7	65	71.4	18	4	US-09-128-345-67
8	63	69.2	18	1	US-08-499-523-53
9	63	69.2	18	1	US-08-499-523-58
10	63	69.2	18	4	US-09-128-345-53
11	63	69.2	18	4	US-09-128-345-58
12	62	68.1	18	1	US-08-499-523-54
13	62	68.1	18	1	US-08-499-523-59
14	62	68.1	18	4	US-09-128-345-54
15	62	68.1	18	4	US-09-128-345-59
16	61	67.0	18	1	US-08-095-769A-1
17	61	67.0	18	1	US-08-182-483A-2
18	61	67.0	18	1	US-08-182-483A-28
19	61	67.0	18	1	US-08-243-879A-1
20	61	67.0	18	1	US-08-243-879A-27
21	61	67.0	18	1	US-08-499-523-11
22	61	67.0	18	1	US-08-499-523-16
23	61	67.0	18	1	US-08-499-523-33
24	61	67.0	18	1	US-08-499-523-48
25	61	67.0	18	2	US-08-752-852A-1
26	61	67.0	18	2	US-08-752-852A-9
27	61	67.0	18	2	US-08-752-852A-123

Sequence 154, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 11, Appl  
Sequence 16, Appl  
Sequence 33, Appl  
Sequence 48, Appl  
Sequence 36, Appl  
Sequence 2, Appl  
Sequence 10, Appl  
Sequence 2, Appl  
Sequence 10, Appl  
Sequence 17, Appl  
Sequence 25, Appl  
Sequence 16, Appl  
Sequence 24, Appl  
Sequence 37, Appl

#### ALIGNMENTS

RESULT 1  
US-08-984-294-6  
; Sequence 6, Application US/08984294  
; Patent No. 6043220  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway C.  
; APPLICANT: Chen, Jie  
; APPLICANT: Lehner, Robert I.  
; TITLE OF INVENTION: THREONINE-CONTAINING PROTEGRINS  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of The Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/984,294  
; FILING DATE: 03-DEC-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 008067-0049-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 6043220e  
; US-08-984-294-6

Query Match 72.5%; Score 66; DB 3; Length 18;  
Best Local Similarity 72.2%; Pred. No. 0.00017;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;



; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 67:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 8, 13, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-67

Query Match 71.4%; Score 65; DB 1; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.00024;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Caps 0;

Qy 1 RGGRLSYRRRFFSTSTGR 18  
| | | | | | | | | |  
Db 1 RGGRLXYRRRFXVXVGR 18

RESULT 5  
US-08-752-852A-230  
; Sequence 230, Application US/08752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrer, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 230:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids

; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-752-852A-230  
  
Query Match 71.4%; Score 65; DB 2; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.00024;  
Matches 12; Conservative 2; Mismatches 4; Indels 0; Caps 0;  
  
Qy 1 RGGRLSYRRRFFSTSTGR 18  
| | | | | | | | | |  
Db 1 RGGRLCYARRRFAVCVGR 18  
| | | | | | | | | |  
  
RESULT 6  
US-09-128-345-63  
; Sequence 63, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HAWIG, SYLVIA S.L.  
; APPLICANT: KORRIAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-0054-999  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 8, 13, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-63

Query Match 71.4%; Score 65; DB 4; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.00024;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Caps 0;  
  
Qy 1 RGGRLSYRRRFFSTSTGR 18  
| | | | | | | | | |  
Db 1 RGGRLXYRRRFXVXVGR 18

```

US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 615936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-09-128-345-67

Query Match 71.4%; Score 65; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00024;
Matches 12; Conservative 0; Mismatches 6; Indels 0;

QY 1 RGGRLSYRRRFFSTGTGR 18
Db 1 RGGRLXYRRRFXVXGR 18
||||| 1 ||||| 11

US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 615936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-09-128-345-67

Query Match 71.4%; Score 65; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00024;
Matches 12; Conservative 0; Mismatches 6; Indels 0;

QY 1 RGGRLSYRRRFFSTGTGR 18
Db 1 RGGRLXYRRRFXVXGR 18
||||| 1 ||||| 11

US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 615936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-09-128-345-67

Query Match 71.4%; Score 65; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00024;
Matches 12; Conservative 0; Mismatches 6; Indels 0;

QY 1 RGGRLSYRRRFFSTGTGR 18
Db 1 RGGRLXYRRRFXVXGR 18
||||| 1 ||||| 11

US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 615936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-09-128-345-67

Query Match 71.4%; Score 65; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00024;
Matches 12; Conservative 0; Mismatches 6; Indels 0;

QY 1 RGGRLSYRRRFFSTGTGR 18
Db 1 RGGRLXYRRRFXVXGR 18
||||| 1 ||||| 11

US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 615936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-09-128-345-67

Query Match 71.4%; Score 65; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00024;
Matches 12; Conservative 0; Mismatches 6; Indels 0;

QY 1 RGGRLSYRRRFFSTGTGR 18
Db 1 RGGRLXYRRRFXVXGR 18
||||| 1 ||||| 11

US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 615936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note="X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
; US-09-128-345-67

Query Match 71.4%; Score 65; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00024;
Matches 12; Conservative 0; Mismatches 6; Indels 0;

QY 1 RGGRLSYRRRFFSTGTGR 18
Db 1 RGGRLXYRRRFXVXGR 18
||||| 1 ||||| 11

US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 615936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER
```

REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
NAME/KEY: Disulfide-bond  
LOCATION: 6..15  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(8, 13)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-58

Query Match 69.2%; Score 63; DB 1; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.00051;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RGGRLYSRRRFSTSTGR 18  
||||| ||||| ||  
Db 1 RGGRLCYRRRFVCVGR 18

## RESULT 10

US-09-128-345-53  
Sequence 53, Application US/09128345  
Patent No. 6159936  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/128, 345  
APPLICATION NUMBER: US/09/128, 345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
NAME/KEY: Disulfide-bond

LOCATION: 8..13  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-53

Query Match 69.2%; Score 63; DB 4; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.00051;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RGGRLYSRRRFSTSTGR 18  
||||| ||||| ||  
Db 1 RGGRLCYRRRFVCVGR 18

## RESULT 11

US-09-128-345-58  
Sequence 58, Application US/09128345  
Patent No. 6159936  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/128, 345  
APPLICATION NUMBER: US/09/128, 345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 6..15  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(8, 13)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-58

Query Match 69.2%; Score 63; DB 4; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.00051;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 1 RGGRLSYRRRSTSTGR 18
      ||||| | |||| | ||
Db 1 RGGRLCYXRRRFXCVGR 18

RESULT 12
US-08-499-523-54
; Sequence 54, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 8..13
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-08-499-523-54

Query Match 68.1%; Score 62; DB 1; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00073;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRSTSTGR 18
      ||||| | |||| | ||
Db 1 RGGRLCYXRRRFXCVGR 18

RESULT 13
US-08-499-523-59
; Sequence 59, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 8..13
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-08-499-523-54

Query Match 68.1%; Score 62; DB 1; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00073;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRSTSTGR 18
      ||||| | |||| | ||
Db 1 RGGRLCYXRRRFXCVGR 18

RESULT 14
US-09-128-345-54
; Sequence 54, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```



```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; NAME/KEY: Disulfide-bond
; LOCATION: 8..13
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 15)
; OTHER INFORMATION: /note= "x is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-54
```

```
Query Match 68.1%; Score 62; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00073;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy 1 RGGRLSYSRRRFSTSTGR 18
    ||||| | ||||
Db 1 RGGRLXCRRRRFCIXVGR 18
```

```
RESULT 15
US-09-128-345-59
; Sequence 59, Application US/09128345
; Patent No. 6159336
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128.345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
```

```
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 6..15
; NAME/KEY: Modified-site
; LOCATION: group(8, 13)
; OTHER INFORMATION: /note= "x is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-59
```

```
Query Match 68.1%; Score 62; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.00073;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy 1 RGGRLSYSRRRFSTSTGR 18
    ||||| | ||||
Db 1 RGGRLCYXRRRFXICVGR 18
```

```
Search completed: February 12, 2002, 12:32:23
Job time: 451 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: February 12, 2002, 12:34:40 ; Search time 126.85 Seconds  
(without alignments)  
10.809 Million cell updates/sec

Title: US-09-485-571-25

Perfect score: 91

Sequence: 1 RGGRLSYRRRSTSTGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	67.0	149	2 S57607	protegrin 1 precur
2	55	60.4	147	2 JN0900	protegrin 2 precur
3	54	59.3	149	2 S57609	protegrin 3 precur
4	54	59.3	149	2 A53895	nucleic acid-bind
5	47	51.6	94	1 B48549	hypothetical prote
6	47	51.6	447	2 S53982	hypothetical prote
7	46	50.5	231	2 T32047	hypothetical prote
8	45	49.5	191	1 O0AG6T	hypothetical prote
9	44	48.4	186	2 A30832	hypothetical prote
10	44	48.4	293	2 D81896	Neisseria mening
11	44	48.4	413	2 T03240	FLO/LFY protein ho
12	44	48.4	532	2 H72730	probable acyl-CoA
13	43	47.3	107	2 J01251	hypothetical 12.6K
14	43	47.3	108	1 W0VVP5	nucleic acid-bind
15	43	47.3	108	2 S12976	12k protein - pota
16	42	46.2	118	2 C72842	hypothetical prote
17	42	46.2	386	2 A41950	retrovirus-related
18	42	46.2	747	2 T42599	minor capsid prote
19	42	46.2	1016	2 T30942	aminopeptidase (EC
20	42	46.2	1016	2 T30943	aminopeptidase (EC
21	41	45.1	201	2 T72739	hypothetical prote
22	41	45.1	310	2 T43147	hypothetical prote
23	41	45.1	325	2 T38308	hypothetical prote
24	41	45.1	547	2 S53920	SNGL protein - yea
25	41	45.1	619	2 A85958	glutathionylspermi
26	41	45.1	619	2 A57338	glutathionylspermi
27	40.5	44.5	464	2 A56600	intermediate filam
28	40	44.0	217	2 T33652	hypothetical prote
29	40	44.0	352	2 D96597	hypothetical prote

30	40	44.0	499	2 C85022	hypothetical prote
31	40	44.0	508	1 O4CHC7	steroid 17alpha-mo
32	40	44.0	513	2 T02002	hypothetical prote
33	40	44.0	622	2 A70414	NADH dehydrogenase
34	40	44.0	622	2 E71384	keratin 9, type I,
35	40	44.0	743	2 E71432	hypothetical prote
36	40	44.0	1030	2 T16114	hypothetical prote
37	40	44.0	1224	2 T26377	hypothetical prote
38	40	44.0	1696	2 T00057	hypothetical prote
39	39	42.9	51	2 G70154	ribosomal protein
40	39	42.9	105	2 S42214	hydroxymethylgluta
41	39	42.9	105	2 S42213	hydroxymethylgluta
42	39	42.9	119	2 D82387	conserved hypother
43	39	42.9	159	2 T05656	hypothetical prote
44	39	42.9	188	2 A48024	glycosylphosphatid
45	39	42.9	211	2 C84888	hypothetical prote

ALIGNMENTS

RESULT 1

S57607

protegrin : precursor - pig

N;Alternate names: neutrophil peptide 1

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 19-Oct-1995 #sequence.revision 03-Nov-1995 #text\_change 16-Jul-1999

C;Accession: S66284; S45712; S36820; S34585; S57607

R;Zhao, C.; Ganz, T.; Lehrer, R.I.

FEBS Lett. 368, 197-202, 1995

A;Title: The structure of porcine protegrin genes.

A;Reference number: S66283; MUID:95354835

A;Accession: S66284

A;Molecule type: DNA

A;Residues: 1-149 <ZHA>

A;Cross-references: EMBL:X84094; NID:g887642; PIDN:CAA58890.1; PID:g887643

R;Zhao, C.; Liu, L.; Lehrer, R.I.

FEBS Lett. 346, 285-288, 1994

A;Title: Identification of a new member of the protegrin family by cDNA cloning.

A;Reference number: S45712; MUID:94283613

A;Accession: S45712

A;Molecule type: mRNA

A;Residues: 1-149 <ZH2>

A;Cross-references: GB:X79868; NID:g603035; PIDN:CAA56251.1; PID:g603036

R;Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Eg

FEBS Lett. 330, 339-342, 1993

A;Title: Primary structure of three cationic peptides from porcine neutrophils. Seq

A;Reference number: S36820; MUID:93387466

A;Accession: S36820

A;Molecule type: protein

A;Residues: 131-148 <MIR>

R;Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M

FEBS Lett. 327, 231-236, 1993

A;Title: Protegrins: leukocyte antimicrobial peptides that combine features of cort

A;Reference number: S34585; MUID:93327946

A;Accession: S34585

A;Molecule type: protein

A;Residues: 131-148 <KOK>

C;Genetics:

A;Gene: NPG.

A;Introns: 66/3; 102/3; 126/3

C;Superfamily: cathelin; cystatin homology

C;Keywords: amidated carboxyl end; antibacterial; neutrophil

F;1-29/Domain: signal sequence #status predicted <SIG>

F;22-129/Domain: cystatin homology <CYS>

F;30-130/Domain: propeptide #status predicted <PRO>

F;131-148/Product: protegrin 1 #status experimental <NAT>

F;148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followi

Query Match 67.0%; Score 61; DB 2; Length 149;

Best Local Similarity 66.7%; Pred. No. 0.0069;

Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

C:Keywords: amidated carboxyl end; antibacterial
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:30-130/Domain: propeptide #status predicted <PRO>
F:131-148/Product: protegrin 5 #status predicted <MAT>
F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followin

Query Match          59.3%; Score 54; DB 2; Length 149;
Best Local Similarity 61.1%; Pred. No. 0.1;
Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFFSTGCR 18
    ||||| | |||| |
Db 131 RGGRLCYCRPRFCVCVGR 148

RESULT 4
A53895
protegrin 3 precursor - pig
N:Alternate names: neutrophil peptide 2
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C:Accession: S66285; A53895; S34587; S36821; S57608
R:Zhao, C.; Ganz, T.; Lehrer, R.I.
FEBS Lett. 368, 197-202, 1995
A:Title: The structure of porcine protegrin genes.
A:Reference number: S66283; MUID:95354835
A:Accession: S66285
A:Molecule type: DNA
A:Residues: 1-149 <ZHA>
A:Cross-references: EMBL:X84095; NID:9887644; PIDN:CAA58891.1; PID:9887645
R:Zhao, C.; Liu, L.; Lehrer, R.I.
FEBS Lett. 346, 285-288, 1994
A:Title: Identification of a new member of the protegrin family by cDNA cloning.
A:Reference number: S45712; MUID:94283613
A:Accession: A53895
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-149 <ZHA>
A:Cross-references: GB:X83267; NID:9603037; PIDN:CAA58240.1; PID:9603038
R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M
FEBS Lett. 327, 231-236, 1993
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of cort.
A:Reference number: S34585; MUID:93327946
A:Accession: S34587
A:Molecule type: protein
A:Residues: 131-148 <KOK>
R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Eg
FEBS Lett. 330, 339-342, 1993
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Seq
A:Reference number: S36820; MUID:93387466
A:Accession: S36821
A:Molecule type: protein
A:Residues: 131-148 <MIR>
C:Genetics:
C:Gene: NPG3
A:Introns: 66/3; 102/3; 126/3
C:Superfamily: cathelin; cystatin homology
C:Keywords: amidated carboxyl end; antibacterial; neutrophil
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:30-130/Domain: propeptide #status predicted <PRO>
F:131-148/Product: protegrin 3 #status experimental <MAT>
F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followin

Query Match          59.3%; Score 54; DB 2; Length 149;
Best Local Similarity 61.1%; Pred. No. 0.1;
Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFFSTGCR 18
    ||||| | |||| |
Db 131 RGGRLCYCRPRFCVCVGR 148

protegrin 2 precursor - pig
N:Alternate names: cathelin-like protein precursor; neutrophil peptide 3
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C:Accession: JN0900; S66282; S34586
R:Storici, P.; Zanetti, M.
Biochem. Biophys. Res. Commun. 196, 1363-1368, 1993
A:Title: A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide with a cat
A:Reference number: JN0900; MUID:94071898
A:Accession: JN0900
A:Molecule type: mRNA
A:Residues: 1-147 <STO>
A:Cross-references: GB:L24745; NID:9431435; PIDN:AAA31061.1; PID:9431436
R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Egorov,
FEBS Lett. 330, 339-342, 1993
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Sequence
A:Reference number: S36820; MUID:93387466
A:Accession: S36822
A:Molecule type: protein
A:Residues: 131-146 <MIR>
R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.; Sh
FEBS Lett. 327, 231-236, 1993
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of corticost
A:Reference number: S34585; MUID:93327946
A:Accession: S34586
A:Molecule type: protein
A:Residues: 131-146 <KOK>
C:Comment: This protein plays a role in the nonoxidative antimicrobial defense mechanism
C:Superfamily: cathelin; cystatin homology
C:Keywords: amidated carboxyl end; antibacterial; neutrophil
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:30-130/Domain: propeptide #status experimental <MAT>
F:131-146/Product: protegrin 2 #status experimental <MAT>
F:146/Modified site: amidated carboxyl end (Val) (amide in mature form from following gl

Query Match          60.4%; Score 55; DB 2; Length 147;
Best Local Similarity 64.7%; Pred. No. 0.067;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFFSTGCR 17
    ||||| | |||| |
Db 131 RGGRLCYCRPRFCVCVGR 147

protegrin 5 precursor - pig
N:Alternate names: cathelin-associated antimicrobial peptide
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 16-Jul-1999
C:Accession: S66283; S57609
R:Zhao, C.; Ganz, T.; Lehrer, R.I.
FEBS Lett. 368, 197-202, 1995
A:Title: The structure of porcine protegrin genes.
A:Reference number: S66283; MUID:95354835
A:Accession: S66283
A:Molecule type: DNA
A:Residues: 1-149 <ZHA>
A:Cross-references: EMBL:X84096; NID:9887646; PIDN:CAA58892.1; PID:9887647
A:Experimental source: leukocytes
C:Genetics:
C:Gene: NPG5
A:Introns: 66/3; 102/3; 126/3
C:Superfamily: cathelin; cystatin homology
```

Db 131 RGGGLCYCRRRRCVCVGR 148

## RESULT 5

B48549

nucleic acid-binding protein - potato virus S

C:Species: potato virus S

C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 23-Jul-1999

C:Accession: B48549

R:Poster, G.D.; Mills, P.R.

Virus Genes 6: 213-220, 1992

A:Title: The 3'-nucleotide sequence of an ordinary strain of potato virus S.

A:Reference number: A48549; MUID:93033173

A:Accession: B48549

A:Molecule type: genomic RNA

A:Residues: 1-94 <FOS>

A:Cross-references: GB:S45593; NID:g256417; PIDN:AAB23462.1; PID:g256419

A>Note: sequence extracted from NCBI backbone (NCBIN:114637, NCBIPI:114639)

C:Superfamily: potato virus nucleic acid-binding protein

C:Keywords: DNA binding; zinc finger

F:55-75/Region: zinc finger

Query Match 51.6%; Score 47; DB 1; Length 94;

Best Local Similarity 52.9%; Pred. No. 0.93;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLSYSRRRFRSTGTGR 18

||||:|:|:|:|

Db 38 GGRSTYARKRRARSIGR 54

## RESULT 6

S53982

hypothetical protein 447 - actinophage RP3

C:Species: actinophage RP3

C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997

C:Accession: S53982

R:Gabriel, K.; Schmid, H.; Schmidt, U.; Rausch, H.

Nucleic Acids Res. 23, 58-63, 1995

A:Title: The actinophage RP3 DNA integrates site-specifically into the putative tRNA(Arg

A:Reference number: S53981; MUID:95175370

A:Accession: S53982

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-447 <GAB>

A:Cross-references: EMBL:X80661

C:Genetics:

A:Start codon: GTG

Query Match 51.6%; Score 47; DB 2; Length 447;

Best Local Similarity 47.1%; Pred. No. 4.1;

Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLYSRRRFRSTGT 17

||||:|:|:|:|

Db 9 RGGSVAYAKRRVSTAKG 25

## RESULT 7

T32047

hypothetical protein K07E8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T32047

R:Jones, K.; Kramer, J.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid K07E8.

A:Reference number: Z21116

A:Accession: T32047

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-231 <JON>

A:Cross-references: EMBL:AF016678; PIDN:AAB66149.1; GSPDB:GN00020; CESP:K07E8.3

A:Experimental source: strain Bristol N2; clone K07E8

C:Genetics:

A:Gene: CESP:K07E8.3

A:Map position: 2

A:Introns: 69/3; 104/1; 171/3; 203/3

Query Match 50.5%; Score 46; DB 2; Length 231;

Best Local Similarity 47.1%; Pred. No. 3.2;

Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLYSRRRFRSTGT 17

||||:|:|:|

Db 74 RGGRDYDHKRYPNRTG 90

## RESULT 8

Q0AG6T

hypothetical protein 6 - Agrobacterium tumefaciens plasmids

C:Species: Agrobacterium tumefaciens

C:Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 16-Jul-1999

C:Accession: A04498; S28691

R:Gielen, J.; De Beuckeleer, M.; Seurinck, J.; Deboeck, F.; De Greve, H.; Lemmers,

EMBO J. 3, 835-846, 1984

A:Title: The complete nucleotide sequence of the TL-DNA of the Agrobacterium tumefaciens

A:Reference number: A91001; MUID:84207942

A:Accession: A04498

A:Molecule type: DNA

A:Residues: 1-191 <GRE>

A:Cross-references: GB:X00493; GB:J05108; GB:X00282; NID:g39062; PIDN:CAA25171.1; P

A:Experimental source: plasmid pTiAch5

R:Barker, R.F.; Idler, K.B.; Thompson, D.V.; Kemp, J.D.

Plant Mol. Biol. 2, 335-350, 1983

A:Title: Nucleotide sequence of the T-DNA region from the Agrobacterium tumefaciens

A:Reference number: S28683

A:Accession: S28691

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-191 <BAR>

A:Cross-references: EMBL:X00493; NID:g39062; PIDN:CAA25171.1; PID:g39071

A:Experimental source: plasmid pTi15955

C:Genetics:

A:Genome: plasmid

C:Superfamily: T-6b protein

C:Keywords: crown gall tumor

Query Match 49.5%; Score 45; DB 1; Length 191;

Best Local Similarity 50.0%; Pred. No. 3.9;

Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGRLSYSRRRFRSTS 15

||||:|:|:|

Db 112 GGRINYSKNEYSSS 125

## RESULT 9

A30832

hypothetical protein 6a - Agrobacterium tumefaciens plasmid pTiT37

C:Species: Agrobacterium tumefaciens

C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 30-Sep-1993

C:Accession: A30832

R:Vanderlinden, J.; Desair, J.; De Meirsmann, C.; Michiels, K.; Van Gool, A.; Jen, G

Plant Mol. Biol. 7, 33-41, 1986

A:Title: Nucleotide sequence of the T-DNA region encoding transcripts 6a and 6b of

A:Reference number: A93763

A:Accession: A30832

A:Molecule type: DNA

A:Residues: 1-186 <VAN>

C:Genetics:

A:Genome: plasmid

Wed Feb 13 07:52:18 2002

us-09-485-571-25.rpr

C:Superfamily: T-6b protein

Query Match 48.4%; Score 44; DB 2; Length 186;

Best Local Similarity 57.1%; Pred. No. 5.6;

Matches 8; Conservative 3; Mismatches 0; Indels 3; Gaps 0;

QY 2 GRLSYRRRFRSTG 15

Db 107 GGRINYSRNEHSS 120

RESULT 10

D81896 Neisseria meningitidis hypothetical protein NMA1282 - Neisseria meningitidis (strain 224

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: D81896

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.

A:Reference number: A81775; MUID:20222556

A:Accession: D81896

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-293 &lt;PAR&gt;

A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84534.1; PID:g737995

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: NMA1282

C:Superfamily: Neisseria meningitidis hypothetical protein NMA1282

Query Match 48.4%; Score 44; DB 2; Length 293;

Best Local Similarity 57.1%; Pred. No. 8.7;

Matches 8; Conservative 2; Mismatches 0; Indels 4; Gaps 0;

QY 4 RLSYSRRRFRSTG 17

Db 38 RLGSRRKNFAEQTG 51

RESULT 11

T03240

FLU/LFY protein homolog NFL1 - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Mar-2001

C:Accession: T03240

R:Kelly, A.J.; Bonlander, M.B.; Meeks-Wagner, D.R.

Plant Cell 7, 225-234, 1995

A:Title: NFL, the tobacco homologue of FLORICAULA and LEAFY, is transcriptionally expres

A:Reference number: 214855; MUID:95276463

A:Accession: T03240

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-413 &lt;KEL&gt;

A:Cross-references: EMBL:U16172; NID:g561681; PIDN:AAC48985.1; PID:g561683

A:Experimental source: cultivar Samsun

C:Genetics:

A:Introns: 154/1; 288/3

C:Superfamily: Arabidopsis thaliana LFY floral meristem identity control protein

C:Keywords: transcription regulation

Query Match 48.4%; Score 44; DB 2; Length 413;

Best Local Similarity 52.9%; Pred. No. 12;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 GRLSYRRRFRSTG 18

Db 186 GGRMKQRKRKVVSTGR 202

RESULT 12

H72730

probable acyl-CoA dehydrogenase, short-chain specific APE0385 - Aeropyrum pernix (st

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C:Accession: H72730

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; T

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aer

A:Reference number: A72450; MUID:99310339

A:Accession: H72730

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-532 &lt;KAW&gt;

A:Cross-references: DBJ:AP000059; NID:g5103911; PIDN:BAA79340.1; PID:dl043126; PID

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0385

Query Match 48.4%; Score 44; DB 2; Length 532;

Best Local Similarity 52.9%; Pred. No. 15;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFRSTG 17

Db 222 RGGGLNYSVRRRLDKSG 238

RESULT 13

JQ1251

hypothetical 12.6K protein - chrysanthemum virus B

C:Species: Chrysanthemum virus B

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Sep-1999

C:Accession: JQ1251

J. Gen. Virol. 72, 2333-2337, 1991

A:Title: Nucleotide sequence and gene organization of the 3'-terminal region of chry

A:Reference number: JQ1246; MUID:92013948

A:Accession: JQ1251

A:Molecule type: genomic RNA

A:Residues: 1-107 &lt;LEV&gt;

A:Cross-references: GB:S60150; NID:g237315; PIDN:AAB20081.1; PID:g237321

C:Superfamily: potato virus nucleic acid-binding protein

Query Match 47.3%; Score 43; DB 2; Length 107;

Best Local Similarity 56.2%; Pred. No. 4.9;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 GRLSYRRRFRSTG 18

Db 40 GRSSYARRRRRALEGR 55

RESULT 14

WMVVP5

nucleic acid-binding protein - potato virus M (strain Russian)

C:Species: potato virus M

A:Note: host Lycopersicon esculentum (tomato)

C:Date: 31-Mar-1990 #sequence\_revision 23-Mar-1995 #text\_change 29-Oct-1999

C:Accession: F54333; PNO006; PNO095; S21606

R:Zavriev, S.K.; Kanyuka, K.V.; Levay, K.E.

J. Gen. Virol. 72, 9-14, 1991

A:Title: The genome organization of potato virus M RNA.

A:Reference number: A54333; MUID:91116326

A:Accession: F54333

A:Molecule type: genomic RNA

A:Residues: 1-108 &lt;ZAV&gt;

A:Cross-references: EMBL:X53062; NID:g61291; PIDN:CAA37237.1; PID:g61297

R:Rupasov, V.V.; Morozov, S.Y.; Kanyuka, K.V.; Zavriev, S.K.

J. Gen. Virol. 70, 1861-1869, 1989  
A:Title: Partial nucleotide sequence of potato virus M RNA shows similarities to potexv  
A:Reference number: A92800; MUID:89293091  
A:Accession: PN0006  
A:Molecule type: mRNA  
A:Residues: 1-79, 'LVSLTMCARNLLMKE' <RUP>  
A:Note: this sequence has been corrected  
R:Zavriev, S.K.; Kanyka, K.V.; Levay, K.E.  
Mol. Biol. (Mosk.) 25, 761-769, 1991  
A:Title: The complete nucleotide sequence of potato virus M genomic RNA.  
A:Reference number: PN0093; MUID:92049299  
A:Accession: PN0095  
A:Molecule type: genomic RNA  
A:Residues: 1-108 <ZA2>  
A:Cross-references: GB:X53062; NID:g61291; PIDN:CAA37237.1; PID:g61297  
A:Note: this is a revision to the sequence from reference A92800  
R:Zavriev, S.K.  
submitted to the EMBL Data Library, May 1990  
A:Reference number: S21601  
A:Accession: S21606  
A>Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-79, 'LVSLTMCARNLLMKE' <ZA3>  
A:Cross-references: EMBL:X53062  
A:Experimental source: Russian wild type  
C:Superfamily: potato virus nucleic acid-binding protein  
C:Keywords: DNA binding; metal binding; nucleotide binding; zinc finger  
F:57-78/Region: zinc finger

Query Match 47.3%; Score 43; DB 1; Length 108;  
Best Local Similarity 52.9%; Pred. No. 4.9;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRSTSTGR 18  
||| |::|:| |  
Db 40 GGRSKYARRRAISAR 56

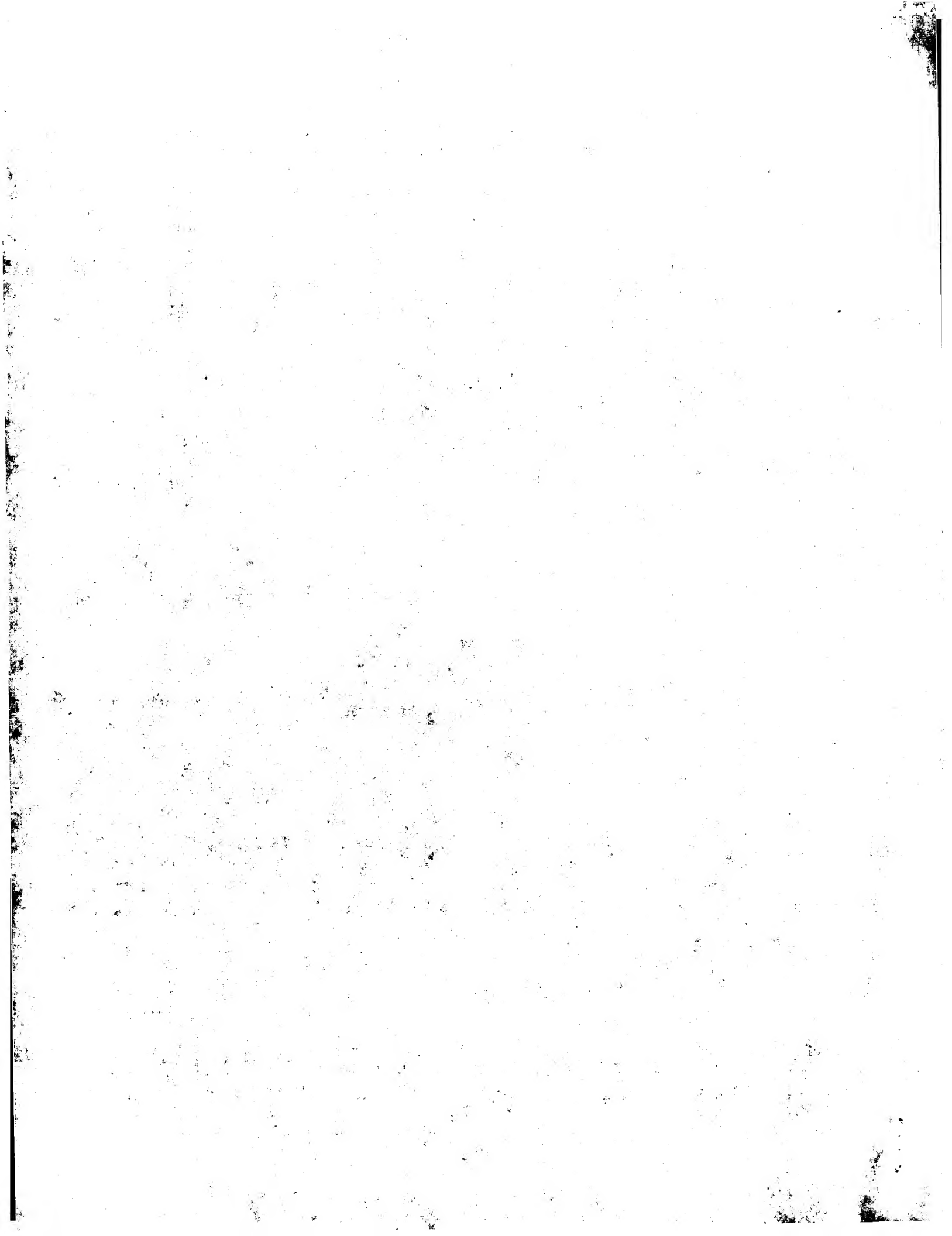
## RESULT 15

S12976  
12K protein - potato virus M  
C:Species: potato virus M  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Sep-1999  
C:Accession: S12976  
R:Gramstat, A.; Courtpozanis, A.; Rohde, W.  
FEBS Lett. 276, 34-38, 1990  
A:Title: The 12 kDa protein of potato virus M displays properties of a nucleic acid-binding  
A:Reference number: S12975; MUID:91092429  
A:Accession: S12976  
A>Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-108 <GRA>  
A:Cross-references: GB:X57440; NID:g61400; PIDN:CAM40689.1; PID:g61402  
C:Superfamily: potato virus nucleic acid-binding protein

Query Match 47.3%; Score 43; DB 2; Length 108;  
Best Local Similarity 52.9%; Pred. No. 4.9;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRSTSTGR 18  
||| |::|:| |  
Db 40 GGRSKYARRRAISAR 56

Search completed: February 12, 2002, 12:34:40  
Job time: 558 sec







RL [6]  
RN STRUCTURE BY NMR OF PROTEGRIN 1.  
RP MEDLINE=97113279; PubMed=8807886;  
RX Fahrner R.L., Dieckmann T., Harwig S.S., Lehrer R.I., Eisenberg D.,  
RA Felgon J.;  
RT "Solution structure of protegrin-1, a broad-spectrum antimicrobial  
peptide from porcine leukocytes.";  
RL Chem. Biol. 3:543-550(1996).  
CC -|- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA  
CC -|- MONOCYTOGENES AND C.ALBIICANS, IN VITRO.  
CC -|- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X79868; CAA56251.1; -;  
DR EMBL; X84094; CAA58890.1; -;  
DR PIR; S34585; S34585.  
DR PIR; S36820; S36820.  
DR PDB; 1PG1; 27-MAY-98.  
DR InterPro; IPR001894; Cathelicidin.  
DR Pfam; PF00666; Cathelicidins; 1.  
DR ProDom; PD001838; Cathelicidin; 1.  
DR PROSITE; PS00946; CATHELICIDINS\_1; 1.  
DR PROSITE; PS00947; CATHELICIDINS\_2; 1.  
KW Antibiotic; Signal; Amidation; Multigene family; 3D-structure.  
FT SIGNAL 1 29 POTENTIAL.  
FT PROPEP 30 130  
FT CHAIN 131 148  
FT MOD\_RES 30 30  
FT PROTEGRIN 1.  
FT PYRROLIDONE CARBOXYLIC ACID (BY  
FT SIMILARITY).  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT MOD\_RES 148 148  
FT SEQUENCE 149 AA; 16677 MW; 6EFBA98429CD6EC4 CRC64;  
AMIDATION (G-149 PROVIDE AMIDE GROUP).  
Query Match 67.0%; Score 61; DB 1; Length 149;  
Best Local Similarity 66.7%; Pred. No. 0.001;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 RGGRLSYRRRSTSTGR 18  
| | | | | | | | | | | | | | | | | | | | | |  
DB 131 RGGRLCYCRRRCVCVGR 148  
RESULT 2  
PG2\_PIG  
ID PG2\_PIG STANDARD; PRT; 147 AA.  
AC P32195;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PROTEGRIN 2 PRECURSOR (PG-2).  
GN NPG2.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=94071898; PubMed=8250892;  
RA Storic P., Zanetti M.;  
RT "A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide  
with a cathelin-like pro-sequence.";

RL Biochem. Biophys. Res. Commun. 196:1363-1368(1993).  
RN [2]  
RP SEQUENCE OF 131-146.  
RX TISSUE=Leukocyte;  
RL MEDLINE=93327946; PubMed=83351113;  
RA Kokryakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,  
RA Aleshina G.M., Shamova O.V., Korneva H.A., Lehrer R.I.;  
RT "Protegrins: leukocyte antimicrobial peptides that combine features  
of corticostatic defenses and tachyplesins.";  
RL FEBS Lett. 327:231-236(1993).  
CC -|- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA  
CC -|- MONOCYTOGENES AND C.ALBIICANS, IN VITRO.  
CC -|- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L24745; AAA31061.1; -;  
DR HSP; P32194; 1PG1  
DR InterPro; IPR001894; Cathelicidin.  
DR Pfam; PF00666; Cathelicidins; 1.  
DR ProDom; PD001838; Cathelicidin; 1.  
DR PROSITE; PS00946; CATHELICIDINS\_1; 1.  
DR PROSITE; PS00947; CATHELICIDINS\_2; 1.  
KW Antibiotic; Signal; Amidation; Multigene family.  
FT SIGNAL 1 29 POTENTIAL.  
FT PROPEP 30 130  
FT CHAIN 131 146  
FT MOD\_RES 30 30  
FT PROTEGRIN 2.  
FT PYRROLIDONE CARBOXYLIC ACID (BY  
FT SIMILARITY).  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT MOD\_RES 146 146  
FT SEQUENCE 147 AA; 16478 MW; 698429DFEC40466 CRC64;  
AMIDATION (G-147 PROVIDE AMIDE GROUP).  
Query Match 60.4%; Score 55; DB 1; Length 147;  
Best Local Similarity 64.7%; Pred. No. 0.011;  
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 RGGRLSYRRRSTSTG 17  
| | | | | | | | | | | | | | | | | | | | | |  
DB 131 RGGRLCYCRRRCVCVG 147  
RESULT 3  
PG3\_PIG  
ID PG3\_PIG STANDARD; PRT; 149 AA.  
AC P32196;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PROTEGRIN 3 PRECURSOR (PG-3).  
GN NPG3.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=94283613; PubMed=8013647;  
RA Zhao C., Liu L., Lehrer R.I.;  
RT "Identification of a new member of the protegrin family by cDNA  
cloning.";  
RL FEBS Lett. 346:285-288(1994).

A;Residues: 1-18 <MI>  
 C;Comment: The peptide is one of the antimicrobial peptides in the American horseshoe crab  
 C;Keywords: amidated carboxyl end  
 F;4-17,8-13/Disulfide bonds: #status predicted  
 F;18/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 52.2% Score 47; DB 2; Length 18;  
 Best Local Similarity 53.8% Pred. No. 0.095;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KWAFRVAYRGIRY 13  
 : I I I I I : I  
 Db 2 RWCRCVCKGFCY 14

RESULT 8

F84162  
 hypothetical protein Vng0026c [imported] - Halobacterium sp. NRC-1  
 C;Species: Halobacterium sp. NRC-1  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C;Accession: F84162  
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A;Title: Genome sequence of Halobacterium species NRC-1.  
 A;Reference number: A84160; MUID:20504483  
 A;Accession: F84162  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-307 <STO>  
 A;Cross-references: GB:AE004437; NID:g10579674; PIDN:AAG18666.1; GSPDB:GN00138  
 C;Genetics:  
 A;Gene: VNG0026C

Query Match 48.9% Score 44; DB 2; Length 307;  
 Best Local Similarity 53.8% Pred. No. 5.8;  
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KWAFRVAYRGIRY 13  
 : I I I I I : I  
 Db 181 KWAFTLYEQVAY 193

RESULT 9

A84161  
 hypothetical protein Vng0013c [imported] - Halobacterium sp. NRC-1  
 C;Species: Halobacterium sp. NRC-1  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C;Accession: A84161  
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A;Title: Genome sequence of Halobacterium species NRC-1.  
 A;Reference number: A84160; MUID:20504483  
 A;Accession: A84161  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-378 <STO>  
 A;Cross-references: GB:AE004437; NID:g10579658; PIDN:AAG18653.1; GSPDB:GN00138  
 C;Genetics:  
 A;Gene: VNG0013C

Query Match 48.9% Score 44; DB 2; Length 378;  
 Best Local Similarity 53.8% Pred. No. 7.1;  
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KWAFRVAYRGIRY 13  
 : I I I I I : I  
 Db 252 KWAFTLYEQVAY 264

RESULT 10

T21986  
 hypothetical protein F39B2.4a - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C;Accession: T21986  
 R;Dobson, R.  
 submitted to the EMBL Data Library, March 1997  
 A;Reference number: Z19498  
 A;Accession: T21986  
 A;Status: preliminary; translated from GB/EMBL/DBDJ  
 A;Molecule type: DNA  
 A;Residues: 1-1661 <WIL>  
 A;Cross-references: EMBL:Z92834; PIDN:CAB07385.1; GSPDB:GN00019; CESP:F39B2.4a  
 A;Experimental source: clone F39B2  
 C;Genetics:  
 A;Gene: CESP:F39B2.4a  
 A;Map position: 1  
 A;Introns: 37/3; 102/2; 229/3; 317/3; 375/2; 505/3; 546/1; 668/3; 833/1; 904/2; 941

Query Match 48.9% Score 44; DB 2; Length 1661;  
 Best Local Similarity 40.9% Pred. No. 33;  
 Matches 9; Conservative 4; Mismatches 3; Indels 6; Gaps 1;

Qy 1 KWAFTLYEQVAY 16  
 : I I : I I I : I  
 Db 147 KWKFINDCIPKIDYKGINILR 168

RESULT 11

T21993  
 hypothetical protein F39B2.4b - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C;Accession: T21993  
 R;Dobson, R.  
 submitted to the EMBL Data Library, March 1997  
 A;Reference number: Z19498  
 A;Accession: T21993  
 A;Status: preliminary; translated from GB/EMBL/DBDJ  
 A;Molecule type: DNA  
 A;Residues: 1-1663 <WIL>  
 A;Cross-references: EMBL:Z92834; PIDN:CAB07394.1; GSPDB:GN00019; CESP:F39B2.4b  
 A;Experimental source: clone F39B2  
 C;Genetics:  
 A;Gene: CESP:F39B2.4b  
 A;Map position: 1  
 A;Introns: 37/3; 102/2; 229/3; 317/3; 375/2; 505/3; 546/1; 670/3; 835/1; 906/2; 943

Query Match 48.9% Score 44; DB 2; Length 1663;  
 Best Local Similarity 40.9% Pred. No. 33;  
 Matches 9; Conservative 4; Mismatches 3; Indels 6; Gaps 1;

Qy 1 KWAFTLYEQVAY 16  
 : I I : I I I : I  
 Db 147 KWKFINDCIPKIDYKGINILR 168

RESULT 12

S76323  
 hypothetical protein sll0310 - Synecocystis sp. (strain PCC 6803)  
 C;Species: Synecocystis sp.  
 A;Variety: PCC 6803  
 C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C;Accession: S76323  
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima

O, K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
S.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S76323  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-302 <KAN>  
A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDN:BAA10175.1; PID:g100154  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 47.8%; Score 43; DB 2; Length 302;

Best Local Similarity 60.0%; Pred. No. 8.4;

Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRYLL 15

Db 116 KWGFRPAYTFRYRL 130

RESULT 13

T43341  
alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (EC 2.4.1.101)  
C:Species: *Caenorhabditis elegans*  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T43341  
R:Chen, S.; Zhou, S.; Sarkar, M.; Spence, A.M.; Schachter, H.  
J. Biol. Chem. 274, 288-297, 1999  
A>Title: Expression of three *Caenorhabditis elegans* N-acetylglucosaminyltransferase I genes  
A:Reference number: Z22441; MUID:99085022

A:Accession: T43341

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-467 <CHE>

A:Cross-references: EMBL:AF082011; NID:g3420841; PIDN:AAD03023.1; PID:g3420842

C:Genetics:

A:Gene: gly-12

A:Map position: X

C:Superfamily: alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 47.8%; Score 43; DB 2; Length 467;

Best Local Similarity 53.8%; Pred. No. 13;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13

Db 153 KWALDKAFKGFY 165

RESULT 14

G69860  
hypothetical protein ykoU - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: G69860  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galle  
A:Authors: Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
A:Authors: Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mance  
Y, M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serod  
A:Authors: Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Togmon, K.; Tosato, V.; Uchiyama  
A:Authors: Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A:Reference number: A69580; MUID:98044033  
A:Accession: G69860  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-611 <KUN>  
A:Cross-references: GB:Z99110; GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13213.1  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ykoU

Query Match 47.8%; Score 43; DB 2; Length 611;

Best Local Similarity 41.2%; Pred. No. 17;

Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRYLLRL 17

Db 18 EMRYEVKDYGRCLIKI 34

RESULT 15

T16401  
alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (EC 2.4.1.101)  
C:Species: *Caenorhabditis elegans*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T16401  
R:Pauley, A.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of *C. elegans* cosmid F48E3.  
A:Reference number: Z18508  
A:Accession: T16401  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-702 <PAU>  
A:Cross-references: EMBL:U28735; NID:g860708; PID:g860709; PIDN:AAA68263.1; CESP:F  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:F48E3.1  
A:Introns: 22/1; 81/3; 130/3; 156/3; 234/3; 290/3; 311/3; 343/2; 382/3; 422/1; 448/2  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 47.8%; Score 43; DB 2; Length 702;

Best Local Similarity 53.8%; Pred. No. 20;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13

Db 402 KWALDKAFKGFY 414

Search completed: February 12, 2002, 12:34:42

Job time: 560 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:54 ; Search time 67.2 Seconds  
(without alignments)  
9.275 Million cell updates/sec

Title: US-09-485-571-32

Perfect score: 90

Sequence: 1 KWAFRVAYRGIRYLRL 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	57	63.3	17	1 TAC1_TACGI	P23684 tachypleus
2	57	63.3	17	1 TAC3_TACGI	P18252 tachypleus
3	57	63.3	77	1 TAC1_TACTR	P14213 tachypleus
4	54	60.0	77	1 TAC2_TACTR	P14214 tachypleus
5	50	55.6	18	1 PPM1_LIMPO	P14215 limulus pol
6	47	52.2	18	1 PPM2_LIMPO	P14216 limulus pol
7	44	48.9	1587	1 SUR2_CABEL	Q10669 caenorhabdi
8	42.5	47.2	752	1 HYSA_PROAC	Q59634 propionibac
9	42	46.7	90	1 IATP_SCHPO	Q74523 schizosacch
10	41	45.6	174	1 CSF3_SHEEP	Q28746 ovis aries
11	41	45.6	195	1 CSF3_BOVIN	P35833 bos taurus
12	40	44.4	173	1 CRGA_HUMAN	P11844 homo sapien
13	40	44.4	304	1 TUA3_AGRVI	P70785 agrobacteri
14	40	44.4	507	1 YCGG_ECOLI	P75995 escherichia
15	40	44.4	529	1 Y632_CHLTR	O84637 chlamydia t
16	40	44.4	529	1 Y921_CHLMU	Q9pjb1 chlamydia m
17	39	43.3	184	1 MPL_MPLV	P40931 myeloprolif
18	39	43.3	516	1 GLPD_MYCTU	Q10502 mycobacteri
19	39	43.3	553	1 GLPD_SINY3	P74257 synecocyst
20	39	43.3	625	1 TPOR_MOUSE	Q08351 mus musculu
21	38	42.2	181	1 WCAF_ECOLI	P71240 escherichia
22	38	42.2	465	1 ARBB_ERWCH	P26206 erwinia chr
23	37.5	41.7	558	1 Y561_CHLMU	Q9pka6 chlamydia m
24	37	41.1	174	1 CRGA_BOVIN	P02527 bos taurus
25	37	41.1	174	1 CRGB_BOVIN	P02526 bos taurus
26	37	41.1	174	1 CRGB_MOUSE	P04344 mus musculu
27	37	41.1	174	1 CRGB_RAT	P10066 rattus norv
28	37	41.1	201	1 MGF_CHICK	Q8uz23 pyrococcus
29	37	41.1	263	1 TRUA_PYRAB	Q8uz23 pyrococcus
30	37	41.1	326	1 AAIR_BOVIN	P28190 bos taurus
31	37	41.1	326	1 AAIR_CANFA	P11616 canis fami
32	37	41.1	326	1 AAIR_CAVPO	P47745 cavia porce
33	37	41.1	326	1 AAIR_HUMAN	P30542 homo sapien

34	37	41.1	326	1	AAIR_RAT	P25099 rattus norv
35	37	41.1	328	1	AAIR_RABIT	P34970 oryctolagus
36	37	41.1	357	1	SFAL_STRFR	Q03424 streptomyce
37	37	41.1	452	1	YD04_SCHPO	O14197 schizosacch
38	37	41.1	551	1	YKZ7_YEAST	P36113 saccharomyc
39	37	41.1	671	1	AMOI_ASPNG	Q12556 aspergillus
40	37	41.1	679	1	NCPR_DROME	Q27597 drosophila
41	37	41.1	1003	1	TP3A_MOUSE	O70157 mus musculu
42	37	41.1	1199	1	NIFJ_SINY3	P52965 synecocyst
43	36.5	40.6	901	1	CR14_MAIZE	O24585 zea mays (m
44	36	40.0	148	1	RS13_PYRHO	O74021 pyrococcus
45	36	40.0	158	1	CRGF_MOUSE	Q03740 mus musculu

ALIGNMENTS

RESULT	TAC1_TACGI	+
ID	TAC1_TACGI	STANDARD; PRT; 17 AA.
AC	P23684;	01-NOV-1991 (Rel. 20, Created)
DT	01-NOV-1991	(Rel. 20, Last sequence update)
DT	01-NOV-1995	(Rel. 32, Last annotation update)
DE	TACHYPLESIN I.	
OS	Tachypleus gigas (Southeast Asian horseshoe crab), and	
OC	Carcinoscorpius rotundicauda (Southeast Asian horseshoe	
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;	
OC	Limulidae; Tachypleus.	
OX	NCBI_TaxID=6852, 6848;	
RN	[1]	
RP	SEQUENCE.	
RC	SPECIES=T. gigas, and C. rotundicauda;	
RX	MEDLINE=91035357; PubMed=2229025;	
RA	Muta T., Fujimoto T., Nakajima H., Iwanaga S.;	
RT	"Tachypleus isolated from hemocytes of Southeast Asian horseshoe	
RT	crabs (Carcinoscorpius rotundicauda and Tachypleus gigas):	
RT	identification of a new tachypleusin, tachypleusin III, and a	
RT	processing intermediate of its precursor.;"	
RL	J. Biochem. 108:261-266(1990).	
CC	-1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.	
DR	PIR; A38824; A38824.	
DR	PIR; JX0124; JX0124.	
KW	Antibiotic; Amidation.	
FT	DISULFID 3 16	
FT	DISULFID 7 12	
FT	MOD_RES 17 17	
SO	SEQUENCE 17 AA; 2269 MW; E9E09BD9D2923C94 CRC64;	
Query Match 63.3%; Score 57; DB 1; Length 17;		
Best Local Similarity 76.9%; Pred. No. 0.00089;		
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
Qy	1 KWAFRVAYRGIRY 13	
Db	1 KWCPRVCYRGICY 13	
RESULT	TAC3_TACGI	+
ID	TAC3_TACGI	STANDARD; PRT; 17 AA.
AC	P18252;	01-NOV-1990 (Rel. 16, Created)
DT	01-NOV-1990	(Rel. 16, Last sequence update)
DT	01-NOV-1995	(Rel. 32, Last annotation update)
DE	TACHYPLESIN III.	
OS	Tachypleus gigas (Southeast Asian horseshoe crab).	
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;	
OC	Limulidae; Tachypleus.	
OX	NCBI_TaxID=6852;	
RN	[1]	
RP	SEQUENCE.	

Wed Feb 13 07:52:40 2002

```

RX MEDLINE=91035357; PubMed=2229025;
RA Muta T., Fujimoto T., Nakajima H., Iwanaga S.;
RT "Tachyplesins isolated from hemocytes of Southeast Asian horseshoe
RT crabs (Carcinoseorus rotundicauda and Tachyplesus gigas);
RT identification of a new tachyplesin, tachyplesin III, and a
RT processing intermediate of its precursor.";
RL J. Biochem. 108:261-266(1990).
CC -1- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND
CC GRAM-POSITIVE BACTERIA.
CC -1- TISSUE SPECIFICITY: HEMOCYTES.
CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.
CC PIR; JX0125; JX0125.
DR Antibiologic; Amidation.
KW DISULFID 3 16
FT DISULFID 7 12
FT MOD_RES 17 17
SQ SEQUENCE 17 AA; 2241 MW; E9E08CE9D2923C94 CRC64;

Query Match 63.3%; Score 57; DB 1; Length 17;
Best Local Similarity 76.9%; Pred. No. 0.00089; 0; Mismatches 3; Indels 0; Gaps 0;
Matches 10; Conservative

QY 1 KWAFRVAYRGIRY 13
DB 1 KWCFRVCYRGICY 13

RESULT 3
TACL_TACTR STANDARD; PRT; 77 AA.
ID TACL_TACTR
AC P14213;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TACHYPLESIN I PRECURSOR.
OS Tachyplesin tridentatus (Japanese horseshoe crab).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Tachyplesus.
OX NCBI_TaxID=6853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91065956; PubMed=2250028;
RA Shigenaga T., Muta T., Toh Y., Tokunaga F., Iwanaga S.;
RT "Antimicrobial tachyplesin peptide precursor. cDNA cloning and
RT cellular localization in the horseshoe crab (Tachyplesus
RT tridentatus).";
RL J. Biol. Chem. 265:21350-21354(1990).
RN [2]
RP SEQUENCE OF 24-40, AND DISULFIDE BONDS.
RX MEDLINE=89034158; PubMed=3141410;
RA Nakamura T., Furunaka H., Miyata T., Tokunaga F., Muta T., Iwanaga S.,
RA Niwa M., Takao T., Shimonishi Y.;
RT "Tachyplesin, a class of antimicrobial peptide from the hemocytes of
RT the horseshoe crab (Tachyplesus tridentatus). Isolation and chemical
RT structure.";
RL J. Biol. Chem. 263:16709-16713(1988).
RN [3]
RP STRUCTURE BY NMR OF 24-40.
RX MEDLINE=90368729; PubMed=2394727;
RA Kawano K., Yoneya T., Miyata T., Yoshikawa K., Tokunaga F.,
RA Terada Y., Iwanaga S.;
RT "Antimicrobial peptide, tachyplesin I, isolated from hemocytes of the
RT horseshoe crab (Tachyplesus tridentatus). NMR determination of the
RT beta-sheet structure.";
RL J. Biol. Chem. 265:15365-15367(1990).
RN [4]
RP STRUCTURE BY NMR OF 24-40.
RX MEDLINE=93257488; PubMed=8490053;
RA Tamamura H., Kuroda M., Masuda M., Otaka A., Funakoshi S.,
RA Nakashima H., Yamamoto N., Waki M., Matsumoto A., Lancellin J.-M.,
RA Kohda D., Tate S., Inagaki F., Fujii N.;
RT "A comparative study of the solution structures of tachyplesin I and
RT

```

```

RT a novel anti-HIV synthetic peptide, T22 ([Tyr5,12, Lys7]-polyphemusin
RT I1), determined by nuclear magnetic resonance.";
RL Biochim. Biophys. Acta 1163:209-216(1993).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=94110249; PubMed=8282718;
RA Shigenaga T., Takayenoki Y., Kawasaki S., Seki N., Muta T., Toh Y.,
RA Ito A., Iwanaga S.;
RT "Separation of large and small granules from horseshoe crab
RT (Tachyplesus tridentatus) hemocytes and characterization of their
RT components.";
RL J. Biochem. 114:307-316(1993).
CC -1- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND
CC GRAM-POSITIVE BACTERIA.
CC -1- SUBCELLULAR LOCATION: S-GRAULES.
CC -1- TISSUE SPECIFICITY: HEMOCYTES.
CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M57242; AAC63538.1; -.
CC PIR; A30068; A30068.
CC PIR; A38345; A38345.
DR Antibiologic; Amidation; Cleavage on pair of basic residues; Signal.
KW SIGNAL 1 23
FT PEPTIDE 24 40 TACHYPLESIN I.
FT PROPEP 41 77
FT DISULFID 26 39
FT DISULFID 30 35
FT MOD_RES 40 40
FT DOMAIN 69 77
FT SEQUENCE 77 AA; 9349 MW; B940CAA4A641335F CRC64;
SQ SEQUENCE 77 AA; 9349 MW; B940CAA4A641335F CRC64;

Query Match 63.3%; Score 57; DB 1; Length 77;
Best Local Similarity 76.9%; Pred. No. 0.0042; 0; Mismatches 3; Indels 0; Gaps 0;
Matches 10; Conservative

QY 1 KWAFRVAYRGIRY 13
DB 24 KWCFRVCYRGICY 36

RESULT 4
TACL_TACTR STANDARD; PRT; 77 AA.
ID TACL_TACTR
AC P14214;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TACHYPLESIN II PRECURSOR.
OS Tachyplesus tridentatus (Japanese horseshoe crab).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Tachyplesus.
OX NCBI_TaxID=6853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91065956; PubMed=2250028;
RA Shigenaga T., Muta T., Toh Y., Tokunaga F., Iwanaga S.;
RT "Antimicrobial tachyplesin peptide precursor. cDNA cloning and
RT cellular localization in the horseshoe crab (Tachyplesus
RT tridentatus).";
RL J. Biol. Chem. 265:21350-21354(1990).
RN [2]
RP SEQUENCE OF 24-40.
RX MEDLINE=90110066; PubMed=2514185;
RA Miyata T., Tokunaga F., Yonega T., Yoshikawa K., Iwanaga S., Niwa M.,

```



RA Takao T., Shimonishi Y.;  
 RT "Antimicrobial peptides, isolated from horseshoe crab hemocytes,  
 RT tachyplesin II, and polyphemusins I and II: chemical structures and  
 RT biological activity.";  
 RL J. Biochem. 106:663-668(1989).  
 RN [3].  
 RN CHARACTERIZATION.  
 RP MEDLINE-94110249; PubMed-828718;  
 RX Shigenaga T., Takayenoki Y., Kawasaki S., Seki N., Muta T., Toh Y.,  
 RA Ito A., Iwanaga S.;  
 RA "Separation of large and small granules from horseshoe crab  
 RT (tachyplesin tridentatus) hemocytes and characterization of their  
 RT components.";  
 RT J. Biochem. 114:307-316(1993).  
 RL J. Biochem. 114:307-316(1993).  
 CC -1- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND  
 CC GRAM-POSITIVE BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: S-GRANULES.  
 CC -1- TISSUE SPECIFICITY: HEMOCYTES.  
 CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
 DR PIR; B38345; B38345.  
 DR PIR; JU0123; JU0123.  
 KW Antibiotic; Amidation; Cleavage on pair of basic residues; Signal.  
 FT SIGNAL 1 23  
 FT PEPTIDE 24 40 TACHYPLESIN II.  
 FT PROPEP 41 77  
 FT DISULFID 26 39 BY SIMILARITY.  
 FT DISULFID 30 35 BY SIMILARITY.  
 FT MOD\_RES 40 40 AMIDATION (G-41 PROVIDE AMIDE GROUP).  
 FT DOMAIN 69 77 ASP/GLU-RICH (ACIDIC).  
 SQ SEQUENCE 77 AA; 9335 MW; 6EBE57A4A652AEFF CRC64;  
 Query Match 60.0%; Score 54; DB 1; Length 77;  
 Best Local Similarity 69.2%; Pred. No. 0.013;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 KWAFRVAYRGIRY 13  
 Db 24 RWCFRVCYRGICY 36  
 RESULT 5  
 ID PPM1\_LIMPO STANDARD; PRT; 18 AA.  
 AC P14215;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE POLYPHEMUSIN I.  
 OS Limulus polyphemus (Atlantic horseshoe crab).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
 OC Limulidae; Limulus.  
 OX NCBI\_TaxID=6850;  
 RN [1]  
 RN SEQUENCE, AND DISULFIDE BONDS.  
 RP MEDLINE-90110066; PubMed-2514185;  
 RX Miyata T., Tokunaga F., Yonega T., Yoshikawa K., Iwanaga S., Niwa M.,  
 RA Takao T., Shimonishi Y.;  
 RA "Antimicrobial peptides, isolated from horseshoe crab hemocytes,  
 RT tachyplesin II, and polyphemusins I and II: chemical structures and  
 RT biological activity.";  
 RL J. Biochem. 106:663-668(1989).  
 CC -1- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND  
 CC GRAM-POSITIVE BACTERIA.  
 CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
 DR PIR; JU0125; JU0125.  
 KW Antibiotic; Amidation.  
 FT DISULFID 4 17 BY SIMILARITY.  
 FT DISULFID 8 13 BY SIMILARITY.  
 FT MOD\_RES 18 18 AMIDATION.  
 SQ SEQUENCE 18 AA; 2431 MW; E402A109D2923504 CRC64;  
 Query Match 52.2%; Score 47; DB 1; Length 18;  
 Best Local Similarity 53.8%; Pred. No. 0.047;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 KWAFRVAYRGIRY 13  
 Db 2 RWCFRVCYKGFCY 14  
 RESULT 7  
 ID SUR2\_CAEL STANDARD; PRT; 1587 AA.  
 AC Q10669;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE SUR-2 PROTEIN.  
 GN SUR-2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=BRISTOL N2;  
 RX MEDLINE-96018822; PubMed-7557379;  
 RA Singh N., Han M.;  
 RT "sur-2, a novel gene, functions late in the let-60 ras-mediated  
 RT signaling pathway during Caenorhabditis elegans vulval induction.";  
 RL Genes Dev. 9:2251-2265(1995).

Query Match 55.6%; Score 50; DB 1; Length 18;  
 Best Local Similarity 61.5%; Pred. No. 0.015;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 KWAFRVAYRGIRY 13  
 Db 2 RWCFRVCYRGFCY 14  
 RESULT 5  
 ID PPM2\_LIMPO STANDARD; PRT; 18 AA.  
 AC P14216;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE POLYPHEMUSIN II.  
 OS Limulus polyphemus (Atlantic horseshoe crab).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
 OC Limulidae; Limulus.  
 OX NCBI\_TaxID=6850;  
 RN [1]  
 RN SEQUENCE.  
 RP MEDLINE-90110066; PubMed-2514185;  
 RX Miyata T., Tokunaga F., Yonega T., Yoshikawa K., Iwanaga S., Niwa M.,  
 RA Takao T., Shimonishi Y.;  
 RA "Antimicrobial peptides, isolated from horseshoe crab hemocytes,  
 RT tachyplesin II, and polyphemusins I and II: chemical structures and  
 RT biological activity.";  
 RL J. Biochem. 106:663-668(1989).  
 CC -1- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND  
 CC GRAM-POSITIVE BACTERIA.  
 CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
 DR PIR; JU0125; JU0125.  
 KW Antibiotic; Amidation.  
 FT DISULFID 4 17 BY SIMILARITY.  
 FT DISULFID 8 13 BY SIMILARITY.  
 FT MOD\_RES 18 18 AMIDATION.  
 SQ SEQUENCE 18 AA; 2431 MW; E402A109D2923504 CRC64;  
 Query Match 52.2%; Score 47; DB 1; Length 18;  
 Best Local Similarity 53.8%; Pred. No. 0.047;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 KWAFRVAYRGIRY 13  
 Db 2 RWCFRVCYKGFCY 14  
 RESULT 7  
 ID SUR2\_CAEL STANDARD; PRT; 1587 AA.  
 AC Q10669;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE SUR-2 PROTEIN.  
 GN SUR-2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=BRISTOL N2;  
 RX MEDLINE-96018822; PubMed-7557379;  
 RA Singh N., Han M.;  
 RT "sur-2, a novel gene, functions late in the let-60 ras-mediated  
 RT signaling pathway during Caenorhabditis elegans vulval induction.";  
 RL Genes Dev. 9:2251-2265(1995).

Wed Feb 13 07:52:40 2002

us-09-485-571-32.isp

CC -!- FUNCTION: FUNCTIONS IN THE LET-60 RAS SIGNALING PATHWAY; ACTS  
 CC DOWNSTREAM OF LET-60 DURING C. ELEGANS VULVAL INDUCTION.  
 CC -!- DEVELOPMENTAL STAGE: HIGHEST LEVELS IN EMBRYOS AND LARVAE.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U33051; AAA85507.1; - ACIDIC.  
 CC DOMAIN 1387 1404  
 CC FT DOMAIN 1405 1587 GLN/HIS-RICH.  
 CC SEQUENCE 1587 AA; 183906 MW; A125FCA74922B11C CRC64;

Query Match 48.9%; Score 44; DB 1; Length 1587;  
 Best Local Similarity 40.9%; Pred. No. 15;  
 Matches 9; Conservative 4; Mismatches 3; Indels 6; Gaps 1;

Qy 1 KWAF-----RVAYRGIRYLLR 16  
 Db 147 KWKFNDCIPKIDYKGINILR 168

RESULT 8  
 HYSA\_PROAC STANDARD; PRT; 752 AA.  
 AC Q59634;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE HYALURONATE LYASE PRECURSOR (EC 4.2.2.1) (HYALURONIDASE) (HYASE).  
 OS Propionibacterium acnes.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Propionibacterineae; Propionibacteriaceae;  
 CC Propionibacterium.  
 CC NCBI\_TaxID=1747;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=97270208; PubMed=9115089;  
 RA Steiner B.M., Romero-Steiner S., Cruce D., George R.;  
 RT "Cloning and sequencing of the hyaluronate lyase gene from  
 RT Propionibacterium acnes";  
 RL Can. J. Microbiol. 43:315-321(1997).  
 CC -!- CATALYTIC ACTIVITY: HYALURONATE - N 3-(4-DEOXY-BETA-D-GLUC-4-  
 CC ENURONOSYL)-N-ACETYL-D-GLUCOSAMINE.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- SUBCELLULAR LOCATION: CELL-ASSOCIATED.  
 CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 8.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U15927; AAA51650.1; -  
 CC InterPro: IPR003159; Lyase\_8.  
 CC Pfam: PF02278; Lyase\_8; 1.  
 CC KW Lyase; Signal.  
 CC FT SIGNAL 1 32  
 CC CHAIN 33 752 HYALURONATE LYASE.  
 CC SEQUENCE 752 AA; 81910 MW; 60D5DCAA691C41A4 CRC64;

Query Match 47.28; Score 42.5; DB 1; Length 752;  
 Best Local Similarity 53.3%; Pred. No. 12; Indels 1; Gaps 1;  
 Matches 8; Conservative 5; Mismatches 1;

Qy 1 KWAFRVAYRGIRYLL 15  
 Db 450 EWAYRTS-QGMRYLL 463

RESULT 9  
 IATP\_SCHPO STANDARD; PRT; 90 AA.  
 ID IATP\_SCHPO  
 AC Q74523;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PUTATIVE ATPASE INHIBITOR, MITOCHONDRIAL PRECURSOR.  
 GN SPCC70.02C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 CC NCBI\_TaxID=4896;  
 CC [1]  
 RN SEQUENCE FROM N.A.

RP STRAIN=972;  
 RC Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;  
 RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RL -!- FUNCTION: FORMS A ONE-TO-ONE COMPLEX WITH ATPASE TO INHIBIT THE  
 CC ENZYME ACTIVITY COMPLETELY (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE ATPASE INHIBITOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AL023794; CAA19352.1; -  
 CC Hypothetical protein; Mitochondrion; Transit peptide.  
 CC TRANSIT 1 90 MITOCHONDRION.  
 CC CHAIN 1 90 PUTATIVE ATPASE INHIBITOR.  
 CC SEQUENCE 90 AA; 10674 MW; F4C4DF852E3E6909 CRC64;

Query Match 46.7%; Score 42; DB 1; Length 90;  
 Best Local Similarity 44.4%; Pred. No. 1.7;  
 Matches 8; Conservative 5; Mismatches 1; Indels 4; Gaps 1;

Qy 1 KWAFR-----VAYRGIRYL 14  
 Db 3 KYCFRKPACISYRGIRFM 20

RESULT 10  
 CSF3\_SHEEP STANDARD; PRT; 174 AA.  
 ID CSF3\_SHEEP  
 AC Q28746;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GRANULOCYTE COLONY-STIMULATING FACTOR (G-CSF).  
 GN CSF3.  
 OS Ovis aries (Sheep).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Caprinae; Ovis.  
 CC NCBI\_TaxID=9940;  
 CC [1]  
 RN SEQUENCE FROM N.A.

RP MEDLINE=95102116; PubMed=7528579;  
 RA O'Brien P.M., Seow H.F., Rothel J.S., Wood P.R.;  
 RT "Cloning and sequencing of an ovine granulocyte colony-stimulating

factor cDNA.";  
 RL DNA Seq. 4:339-342(1994).  
 CC -1- FUNCTION: GRANULOCYTE/MACROPHAGE COLONY-STIMULATING FACTORS ARE  
 CC CYTOKINES THAT ACT IN HEMATOPOIESIS BY CONTROLLING THE PRODUCTION,  
 CC DIFFERENTIATION, AND FUNCTION OF 2 RELATED WHITE CELL POPULATIONS  
 CC OF THE BLOOD, THE GRANULOCYTES AND THE MONOCYTES-MACROPHAGES. THIS  
 CC CSF INDUCES GRANULOCYTES (BY SIMILARITY).  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- PTM: O-GLYCOSYLATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE IL-6 SUPERFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L07939; AAA68006.1; -;  
 DR HSP: P35833; LBGC.  
 DR InterPro: IPR003573; IL6\_MGF\_GCSF.  
 DR Pfam: PF00489; IL6; 1.  
 DR SMART; SM00126; IL6; 1.  
 DR PROSITE; PS00254; INTERLEUKIN\_6; 1.  
 KW Cytokine; Growth factor; Glycoprotein.  
 FT DISULFID 36 42 BY SIMILARITY.  
 FT DISULFID 64 74 BY SIMILARITY.  
 FT CARBOHYD 133 133 O-LINKED (GALNAC... ) (BY SIMILARITY).  
 SQ SEQUENCE 174 AA; 18806 MW; BA5AA8F8D23ACD1E CRC64;  
 -----  
 Query Match 45.6%; Score 41; DB 1; Length 174;  
 Best Local Similarity 77.8%; Pred. No. 5;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 6 VAYRGIRYL 14  
 :|||||  
 Db 163 LAYRGLRYL 171  
 -----  
 RESULT 11  
 ID CSF3\_BOVIN STANDARD; PRT; 195 AA.  
 AC P35833; Q9TV89;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GRANULOCYTE COLONY-STIMULATING FACTOR PRECURSOR (G-CSF).  
 GN CSF3 OR GCSF.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HOLSTEIN.  
 RA Heidari M., Kehrli M.E. Jr.;  
 RT "Cloning, sequencing, and analysis of cDNA encoding bovine granulocyte  
 RT colony stimulating factor.";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
 RX MEDLINE=94076341; PubMed=7504736;  
 RA Lovejoy B., Cascio D., Eisenberg D.;  
 RT "Crystal structure of canine and bovine granulocyte-colony  
 RT stimulating factor (G-CSF)."  
 RL J. Mol. Biol. 234:640-653(1993).  
 CC -1- FUNCTION: GRANULOCYTE/MACROPHAGE COLONY-STIMULATING FACTORS ARE  
 CC CYTOKINES THAT ACT IN HEMATOPOIESIS BY CONTROLLING THE PRODUCTION,  
 CC DIFFERENTIATION, AND FUNCTION OF 2 RELATED WHITE CELL POPULATIONS  
 CC OF THE BLOOD, THE GRANULOCYTES AND THE MONOCYTES-MACROPHAGES. THIS

CSF INDUCES GRANULOCYTES.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- PTM: O-GLYCOSYLATED.  
 CC -1- SIMILARITY: BELONGS TO THE IL-6 SUPERFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF092533; AAD16102.1; -;  
 DR PDB; LBGC; 31-OCT-93.  
 DR InterPro: IPR003573; IL6\_MGF\_GCSF.  
 DR Pfam: PF00489; IL6; 1.  
 DR PRINTS; PR00433; IL6GCSFMGF.  
 DR ProDom; PD008388; GCSF\_MGF; 1.  
 DR SMART; SM00126; IL6; 1.  
 DR PROSITE; PS00254; INTERLEUKIN\_6; 1.  
 KW Cytokine; Growth factor; Glycoprotein; Signal; 3D-structure.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 195 GRANULOCYTE COLONY-STIMULATING FACTOR.  
 FT DISULFID 57 63  
 FT DISULFID 85 95  
 FT CARBOHYD 154 154 O-LINKED (GALNAC... ) (BY SIMILARITY).  
 FT CONFLICT 93 94 TS -> RG (IN REF. 2).  
 FT HELIX 32 60  
 FT HELIX 65 69  
 FT TURN 70 71  
 FT HELIX 72 75  
 FT TURN 76 76  
 FT HELIX 83 85  
 FT TURN 87 89  
 FT HELIX 92 112  
 FT TURN 113 115  
 FT TURN 118 120  
 FT HELIX 121 145  
 FT HELIX 164 191  
 FT TURN 192 192  
 SQ SEQUENCE 195 AA; 21431 MW; 8C06119E4ADFBA73 CRC64;  
 -----  
 Query Match 45.6%; Score 41; DB 1; Length 195;  
 Best Local Similarity 77.8%; Pred. No. 5.6;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 6 VAYRGIRYL 14  
 :|||||  
 Db 184 LAYRGLRYL 192  
 -----  
 RESULT 12  
 ID CRGA\_HUMAN STANDARD; PRT; 173 AA.  
 AC P11844;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE GAMMA CRYSTALLIN A (GAMMA CRYSTALLIN 5).  
 GN CRGA OR CRYG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88038805; PubMed=3670288;  
 RA Weakin S.O., Du R.P., Tsui L.-C., Breitman M.L.;  
 RT "Gamma-crystallins of the human eye lens: expression analysis of five  
 RT members of the gene family.";  
 RL Mol. Cell. Biol. 7:2671-2679(1987).

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32375; AAB61621.1; -  
CC InterPro; IPR000847; HTH\_LYSR.  
CC Pfam; PF00126; HTH\_1; 1.  
CC PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
CC Transcription regulation; DNA-binding; Plasmid.  
CC DNA\_BIND 18 37 H-T-H MOTIF (BY SIMILARITY).  
CC SEQUENCE 304 AA; 33903 MW; D75979BDD919588F CRC64;  
CC  
CC Query Match 44.4%; Score 40; DB 1; Length 304;  
CC Best Local Similarity 42.9%; Pred. No. 13;  
CC Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
CC  
CC QY 1 KWAFRVAYRGIRYL 14  
CC Db 245 QMTARLAVSGVRVY 258  
CC  
CC RESULT 14  
CC YCGG\_ECOLI  
CC ID YCGG\_ECOLI STANDARD; PRT; 507 AA.  
CC AC P75995;  
CC DT 20-AUG-2001 (Rel. 40, Created)  
CC DT 20-AUG-2001 (Rel. 40, Last sequence update)  
CC DE 20-AUG-2001 (Rel. 40, Last annotation update)  
CC DE HYPOTHETICAL PROTEIN YCGG.  
CC GN YCGG OR B1168  
CC OS Escherichia coli.  
CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
CC OC Escherichia.  
CC OX NCBI\_TaxID=562;  
CC [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN-K12 / MG1655;  
CC RX MEDLINE-97426617; PubMed-9278503;  
CC RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
CC RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
CC RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
CC RA Mau B., Shao Y.;  
CC RT "The complete genome sequence of Escherichia coli K-12.";  
CC RL Science 277:1453-1474(1997).  
CC RN [2]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN-K12;  
CC RX MEDLINE-97061202; PubMed-8905232;  
CC RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
CC RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
CC RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
CC RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
CC RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
CC RA Yano M., Horiuchi T.;  
CC RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
CC corresponding to the 12.7-28.0 min region on the linkage map.";  
CC RL DNA Res. 3:137-155(1996).  
CC [1]  
CC -!- SIMILARITY: CONTAINS 1 DUF2 DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; A8000215; AAC74252.1; ALT\_INIT.  
CC EMBL; D90750; BAA36000.1; ALT\_INIT.  
CC EMBL; D90751; BAA36004.1; ALT\_INIT.  
CC EcoGene; EGI3888; ycgG.  
CC

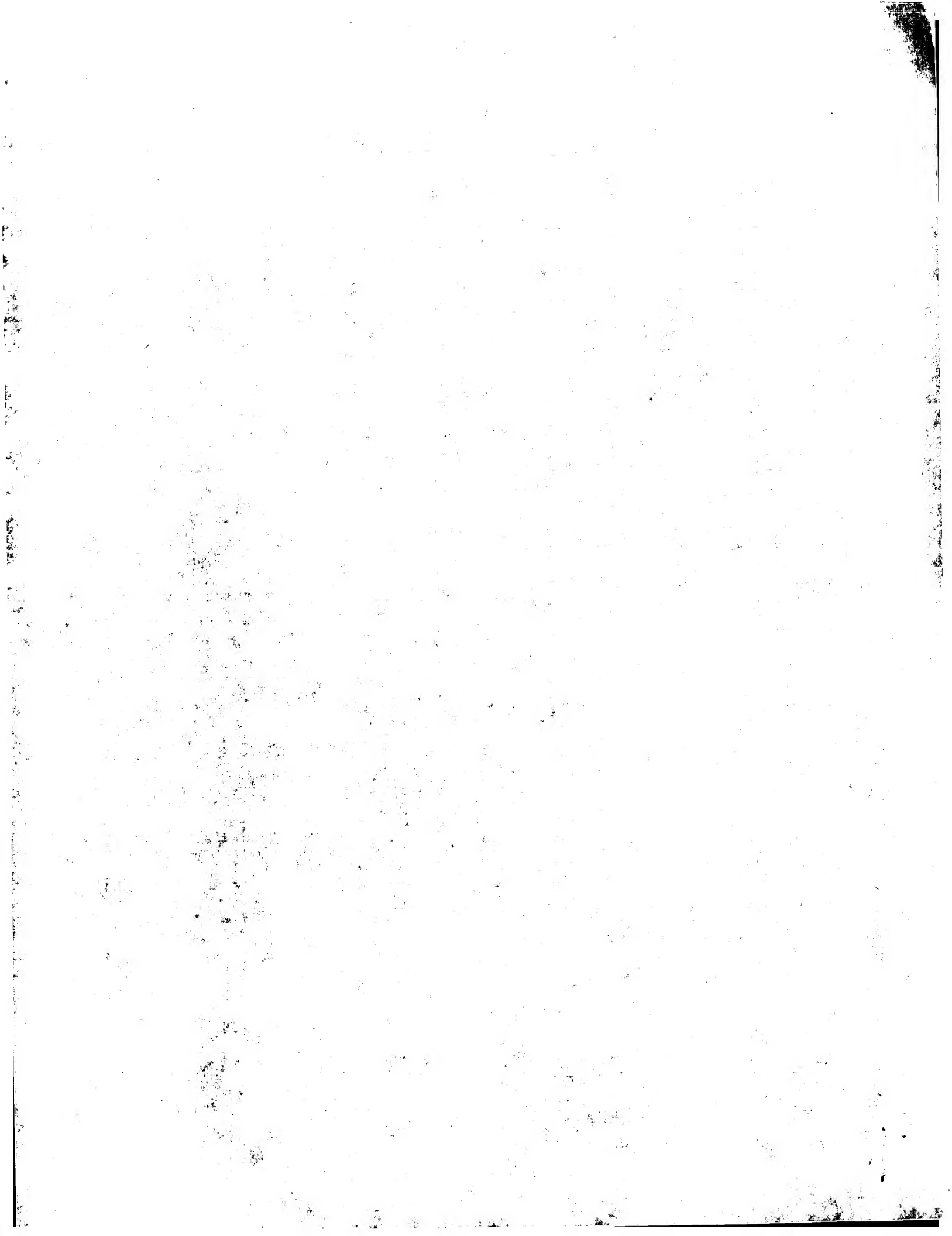
CC -!- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS.  
CC OF THE VITREOUS BODY LENS.  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR  
CC VERY SIMILAR GREEK KEY MOTIFS.  
CC -!- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M17316; AAA52108.1; -  
CC EMBL; M17315; AAA52108.1; JOINED.  
CC PIR; A26912; A26912.  
CC HSP; P02526; IGS.  
CC MIM; 123660; -  
CC InterPro; IPR001064; Crystallin.  
CC Pfam; PF00030; Crystallin; 2.  
CC SMART; SM00247; XTALBQ; 2.  
CC PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 4.  
CC Eye lens protein; Multigene family; Duplication.  
CC INIT\_MET 0 0  
CC DOMAIN 1 39 MOTIF 1.  
CC DOMAIN 40 83 MOTIF 2.  
CC DOMAIN 84 86 CONNECTING PEPTIDE.  
CC DOMAIN 87 127 MOTIF 3.  
CC DOMAIN 128 173 MOTIF 4.  
CC SEQUENCE 173 AA; 20761 MW; 2B4325FEC0E7B070 CRC64;  
CC  
CC Query Match 44.4%; Score 40; DB 1; Length 173;  
CC Best Local Similarity 80.0%; Pred. No. 7.4;  
CC Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
CC  
CC QY 8 YRGIRYLLRL 17  
CC Db 138 YRGIRYLLRL 147  
CC  
CC RESULT 13  
CC TUA3\_AGRVI  
CC ID TUA3\_AGRVI STANDARD; PRT; 304 AA.  
CC AC P70785;  
CC DT 15-DEC-1998 (Rel. 37, Created)  
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
CC DE PROBABLE TARTRATE UTILIZATION TRANSCRIPTIONAL REGULATOR.  
CC TTUA.  
CC GN Agrobacterium vitis.  
CC OS Agrobacterium vitis.  
CC OG Plasmid pTUA3.  
CC OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
CC OC Rhizobiaceae; Rhizobium.  
CC OX NCBI\_TaxID=373;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN-AB3.  
CC RX MEDLINE-96252899; PubMed-8672817;  
CC RA Salomone J.-Y., Crouzet P., de Ruffray P., Otten L.;  
CC RT "Characterization and distribution of tartrate utilization genes in  
CC the grapevine pathogen Agrobacterium vitis.";  
CC RL Mol. Plant Microbe Interact. 9:401-408(1996).  
CC -!- FUNCTION: TRANSCRIPTIONAL REGULATOR OF THE TTUABCODE TARTRATE  
CC UTILIZATION OPERON.  
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

DR InterPro: IPR001633; DUF2.  
DR Pfam: PF00563; DUF2; 1.  
DR SMART: SM00052; DUF2; 1.  
KW Hypothetical protein; Complete proteome.  
FT DOMAIN 245 492 DUF2  
SQ SEQUENCE 507 AA; 56905 MW; D14600B361285D8 CRC64;

Query Match 44.4%; Score 40; DB 1; Length 507;  
Best Local Similarity 46.7%; Pred. No. 22;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
Qy 3 AFRVAYRGIRYLLRL 17  
| | | | | | | | | | | | | | | | | | | | | |  
Db 157 ALNVPLKGVRYVLRV 171

RESULT 15  
Y632\_CHLTR STANDARD; PRT; 529 AA.  
AC O84637;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN CT632.  
GN CT632.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/CX;  
RX MEDLINE=9900809; PubMed=9784136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis.";  
RL Science 282:754-759(1998).  
CC -1- SIMILARITY: BELONGS TO THE UPF0159 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE001334; AAC68236.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 529 AA; 60915 MW; 35BFF03D77AD2D1E CRC64;

Query Match 44.4%; Score 40; DB 1; Length 529;  
Best Local Similarity 35.3%; Pred. No. 23;  
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 KWARFVAYRGIRYLLRL 17  
: | | | | | | | | | | | | | | | | | | | | | |  
Db 460 RWFHINARGLOWLCEL 476



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:43 ; Search time 232.64 Seconds  
(without alignments)  
10.689 Million cell updates/sec

Title: US-09-485-571-32  
Perfect score: 90  
Sequence: 1 KWAFRVAYRGIRYLLRL 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_17:\*
- 1: sp-archaea:\*
  - 2: sp-bacteria:\*
  - 3: sp-fungi:\*
  - 4: sp-human:\*
  - 5: sp-invertebrate:\*
  - 6: sp-mammal:\*
  - 7: sp-mhc:\*
  - 8: sp-organelle:\*
  - 9: sp-phase:\*
  - 10: sp-plant:\*
  - 11: sp-rodent:\*
  - 12: sp-virus:\*
  - 13: sp-vertebrate:\*
  - 14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	53.3	724	2 Q9AD25	Q9ad25 streptomyc
2	45	50.0	513	11 Q9EPH1	Q9eph1 rattus norv
3	44	48.9	307	1 Q9HSY4	Q9hsy4 halobacteri
4	44	48.9	378	1 Q9HSZ6	Q9hsz6 halobacteri
5	44	48.9	1661	5 Q45497	Q45497 caenorhabdi
6	44	48.9	1663	5 Q9U3G5	Q9u3g5 caenorhabdi
7	43	47.8	302	2 Q55655	Q55655 synecocyst
8	43	47.8	467	5 Q76777	Q76777 c udp-n-ace
9	43	47.8	471	5 Q9GPA3	Q9gpa3 caenorhabdi
10	43	47.8	570	10 Q9LQE3	Q9lqe3 arabidopsis
11	43	47.8	611	2 Q34398	Q34398 bacillus su
12	42	46.7	276	2 Q06962	Q06962 salmonella
13	41	45.6	285	2 P73419	P73419 synecocyst
14	41	45.6	323	5 Q9VAQ8	Q9vac8 drosophila
15	41	45.6	475	2 Q9X564	Q9x564 enterococcu
16	41	45.6	976	2 Q9A531	Q9a531 caulobacter
17	40.5	45.0	1500	5 Q9VLQ8	Q9vlq8 drosophila
18	40	44.4	177	5 Q9BL12	Q9bl12 caenorhabdi
19	40	44.4	258	2 Q56836	Q56836 amoeba prot

20	40	44.4	441	8	O20533	O20533 ascosphaera
21	40	44.4	767	10	Q9LQE8	Q9lqe8 arabidopsis
22	40	44.4	789	12	Q9YXE5	Q9yxe5 sphaeopsis
23	40	44.4	925	5	Q9VXK4	Q9vxk4 drosophila
24	40	44.4	949	5	Q9GPJ1	Q9gpj1 heterodera
25	40	44.4	954	5	Q97180	Q97180 drosophila
26	40	44.4	1162	5	Q24818	Q24818 entamoeba h
27	40	44.4	1165	5	Q24860	Q24860 entamoeba h
28	40	44.4	1174	5	Q9GT37	Q9gt37 heterodera
29	40	44.4	1802	6	Q28633	Q28633 oryctolagus
30	40	44.4	2000	6	Q97791	Q97791 oryctolagus
31	40	44.4	26926	4	Q10466	Q10466 homo sapien
32	39.5	43.9	665	2	Q9KNA2	Q9kna2 vibrio chol
33	39.5	43.9	975	5	Q9VN32	Q9vn32 drosophila
34	39	43.3	225	2	Q50898	Q50898 myxococcus
35	39	43.3	267	2	Q9LAU4	Q9lau4 streptomyc
36	39	43.3	298	5	Q21991	Q21991 caenorhabdi
37	39	43.3	300	1	Q29449	Q29449 archaeoglob
38	39	43.3	307	5	Q25546	Q25546 naegleria f
39	39	43.3	357	5	O01564	O01564 caenorhabdi
40	39	43.3	563	5	Q9VRI8	Q9vri8 drosophila
41	39	43.3	1085	2	Q9RJU1	Q9rju1 streptomyc
42	39	43.3	1152	10	Q9C8F4	Q9c8f4 arabidopsis
43	39	43.3	1232	2	O06559	O06559 mycobacteri
44	38.5	42.8	571	5	Q9NA87	Q9na87 caenorhabdi
45	38.5	42.8	868	10	Q38710	Q38710 abies grand

ALIGNMENTS

RESULT	ID	Q9AD25	PRELIMINARY;	PRT;	724 AA.
Q9AD25	AC	Q9AD25;			
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DE	HYPOTHETICAL	79.7 KDA PROTEIN.			
GN	SCPI.125				
OS	Streptomyces coelicolor.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
OX	NCBI_TaxID=1902;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	Brown S.P., Murphy L.D., Harris D.;				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;				
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=A3(2);				
RA	MEDLINE=98241550; PubMed=9573173;				
RT	"Cloning and physical mapping of the EcoRI fragments of the giant linear plasmid SCPI."				
RT	J. Bacteriol. 180:2796-2799(1998).				
RL	EMBL: AL590463; CAC36646.1; -				
DR	Hypothetical protein.				
KW	SEQUENCE 724 AA; 79705 MW; F16B575DF39F77A5 CRC64;				

Query Match 53.3%; Score 48; DB 2; Length 724;  
Best Local Similarity 53.3%; Pred. No. 12;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KWAFRVAYRGIRYLL 15  
:||||:|:|:|

Db 541 RWAFLAARAVRSIL 555

RESULT 2

Q9EPH1 PRELIMINARY; PRT; 513 AA.

AC Q9EPH1; 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 17, Last annotation update)

DE PUTATIVE ALPHA 1B-GLYCOPROTEIN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE DAWLEY; TISSUE=LIVER;

RA Gardmo C., Mode A.;

RT "Cloning of a novel GH regulated rat cDNA homologous to the human

RT alpha1b-glycoprotein.";

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX

CC DOMAIN.

DR EMBL; AJ302031; CAC19029.1; -.

DR InterPro; IPR003599; Ig.

DR InterPro; IPR003600; Ig\_Like.

DR InterPro; IPR003006; Ig\_MHC.

DR Pfam; PF00047; Ig\_2.

DR SMART; SM00409; Ig\_3.

DR SMART; SM00410; Ig\_Like; 1. N -> D.

FT VARIANT 24 24

SQ SEQUENCE 513 AA; 56478 MW; 0991BF67031DAA83 CRC64;

Query Match 50.0%; Score 45; DB 11; Length 513;

Best Local Similarity 75.0%; Pred. No. 26;

Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 RVAYRGIRYLRL 16

Db 154 RVAMRGVTYLLR 165

RESULT 3

Q9HSY4 PRELIMINARY; PRT; 307 AA.

AC Q9HSY4; 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 17, Last annotation update)

DE VNG0026C.

GN VNG0026C.

OS Halobacterium sp. (strain NRC-1).

OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;

OC Halobacterium.

OX NCBI\_TaxID=64091;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE-20504483; PubMed-11016950;

RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,

RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,

RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,

RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,

RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,

RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,

RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,

RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;

RT "Genome sequence of Halobacterium species NRC-1.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

KW EMBL; AE004971; AAG18666.1; -.

KW Complete proteome.

SQ SEQUENCE 307 AA; 34492 MW; AB92A26FF80AF39 CRC64;

Query Match 48.9%; Score 44; DB 1; Length 378;

Best Local Similarity 53.8%; Pred. No. 27;

Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRY 13

Db 252 KWAFRTLYEQVAY 264

RESULT 5

O45497 PRELIMINARY; PRT; 1661 AA.

AC O45497; 06, Created)

DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)

DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)

DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)

DE SUR-2 PROTEIN.

GN SUR-2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Feloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Dobson R.;

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RL MEDLINE-94150718; PubMed-7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,



```

RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL Nature 368:32-38(1994).
DR EMBL: Z92834; CAB07385.1; -.
SQ SEQUENCE 1661 AA; 192896 MW; 36AACA0C7CD86F81 CRC64;

Query Match 48.9%; Score 44; DB 5; Length 1661;
Best Local Similarity 40.9%; Pred. No. 1.4e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 6; Gaps 1;

QY 1 KWAF-----RVAYRGIRYLLR 16
II I :: I:III :II
Db 147 KWKFINDCIPKIDYKGIRNLR 168

RESULT 6
ID Q9U3G5 PRELIMINARY; PRT; 1663 AA.
AC Q9U3G5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE F39B2.4B PROTEIN.
GN F39B2.4B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z92834; CAB07394.1; -.
SQ SEQUENCE 1663 AA; 193108 MW; E2FD5A4D2D6FAA23 CRC64;

Query Match 48.9%; Score 44; DB 5; Length 1663;
Best Local Similarity 40.9%; Pred. No. 1.4e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 6; Gaps 1;

QY 1 KWAF-----RVAYRGIRYLLR 16
II I :: I:III :II
Db 147 KWKFINDCIPKIDYKGIRNLR 168

RESULT 7
ID Q55655 PRELIMINARY; PRT; 302 AA.
AC Q55655;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 33.8 KDA PROTEIN.
GN SLU0310.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=PCC6803;
RA Tabata S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D64000; BAA10175.1; -.
DR InterPro: IPR000182; Acetyltransf_GCN5.
DR Pfam: PF00583; Acetyltransf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 302 AA; 33760 MW; 62D74D7F5B74564E CRC64;

Query Match 47.8%; Score 43; DB 2; Length 302;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIRYLL 15
II III II II I
Db 116 KWGRPAYYTRYRL 130

RESULT 8
ID Q76777 PRELIMINARY; PRT; 467 AA.
AC Q76777;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE-A-3-D-MANNOSIDE B-1,2-N-
DE ACETYLGLUCOSAMINYLTRANSFERASE I (EC 2.4.1.101) (ALPHA-1,3-MANNOSYL-
DE GLYCOPROTEIN BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE) (N-GLYCOSYL-
DE OIGOSACCHARIDE--GLYCOPROTEIN N-ACETYLGLUCOSAMINYLTRANSFERASE I).
GN GLY-12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen S., Zhou S., Sarkar M., Spence A., Schachter H.;
RT "Expression of three Caenorhabditis elegans N-
RT acetylglucosaminyltransferase I genes during development.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + ALPHA-D-MANNOSYL-
CC 1,3-(R1)-BETA-D-MANNOSYL-R2 = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-
CC 1,2-ALPHA-D-MANNOSYL-1,3-(R1)-BETA-D-MANNOSYL-R2.
DR EMBL: AF082011; AAD03023.1; -.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 467 AA; 54363 MW; 02585159B0F9B21D CRC64;

Query Match 47.8%; Score 43; DB 5; Length 467;

```

[illegible]



RA Soier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Yao Q.A.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003768; AAF56843.1; -.  
DR FlyBase; FBgn0039625; CG14517.  
DR InterPro; IPR003859; Galactosyl\_T\_2.  
DR Pfam; PF02709; Galactosyl\_T\_2; 1.  
SQ SEQUENCE 323 AA; 37826 MW; A70251F8888A9DBC CRC64;

Query Match 45.6%; Score 41; DB 5; Length 323;  
Best Local Similarity 60.0%; Pred. No. 72;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 WAFRVAYRGI 11  
Db 208 WRFLPYRGL 217  
| | | | |

RESULT 15  
O9X564 PRELIMINARY; PRT; 475 AA.  
AC O9X564;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE PHOSPHO-BETA-GLUCOSIDASE BGLB.  
GN BGLB.  
OS Enterococcus faecium (Streptococcus faecium).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
OC Enterococcus.  
OX NCBI\_TaxID=1352;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=BFE 900;  
RX MEDLINE=99240446; PubMed=10224016;  
RA Franz C.M.A.P., Worobo R.W., Quadri L.E.N., Schillinger U.,  
RA Holzapfel W.H., Vederas J.C., Stiles M.E.;  
RT "Atypical genetic locus associated with constitutive production of  
enterocin B by enterococcus faecium BFE 900.";  
RL Appl. Environ. Microbiol. 65:2170-2178(1999).  
DR EMBL; AF121254; AAD28227.1; -.  
DR HSSP; P11546; lPEG.  
DR InterPro; IPR001360; Glyco\_hydro\_1.  
DR Pfam; PF00232; Glyco\_hydro\_1; 2.  
DR PRINTS; PR00131; GLHYDRLASE1.  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; UNKNOWN\_1.  
DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
SQ SEQUENCE 475 AA; 54617 MW; D36A3EA8CBF8F12A CRC64;

Query Match 45.6%; Score 41; DB 2; Length 475;  
Best Local Similarity 46.7%; Pred. No. 11e+02;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KWAFRVAYRGIYLL 15  
Db 344 KWGTIDPTGLRYLL 358  
| | : : | : | | |

Search completed: February 12, 2002, 12:38:44  
Job time: 757 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:34 ; Search time 242.57 Seconds  
(without alignments)  
5.191 Million cell updates/sec

Title: US-09-485-571-33  
Perfect score: 93  
Sequence: 1 KYAWRVAHGRIRWLLRX 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	98.9	17	20	AAW99417 Tachyplesin deriva
2	62	66.7	17	20	AAW99416 Tachyplesin deriva
3	45	48.4	20	19	AAW47706 Antimicrobial pept
4	45	48.4	20	20	AAW32548 Antimicrobial pept
5	45	48.4	27	19	AAW47695 Antimicrobial pept
6	45	48.4	27	20	AAW32629 Antimicrobial pept
7	45	48.4	337	22	AAW41298 Human polypeptide
8	44.5	47.8	202	21	AAW44510 Plant viral move
9	44	47.3	27	19	AAW47690 Antimicrobial pept
10	44	47.3	27	20	AAW32624 Antimicrobial pept
11	44	47.3	125	22	AAW95059 Human protein sequ

12	44	47.3	298	21	AAW84945 A murine cholester
13	43	46.2	17	16	AAW75806 Antimicrobial tach
14	43	46.2	17	21	AAW69610 Tachyplesin analog
15	42	45.2	20	19	AAW47701 Antimicrobial pept
16	42	45.2	20	19	AAW47703 Antimicrobial pept
17	42	45.2	20	19	AAW32636 Antimicrobial pept
18	42	45.2	20	20	AAW32547 Antimicrobial pept
19	42	45.2	21	19	AAW47698 Antimicrobial pept
20	42	45.2	21	20	AAW32632 Antimicrobial pept
21	42	45.2	27	19	AAW47689 Antimicrobial pept
22	42	45.2	27	20	AAW32623 Antimicrobial pept
23	42	45.2	28	19	AAW47685 Antimicrobial pept
24	42	45.2	28	20	AAW32619 Antimicrobial pept
25	42	45.2	129	20	AAW59880 Human normal uteru
26	41.5	44.6	508	20	AAW40507 Human lalpha hydro
27	41.5	44.6	508	20	AAW92995 Human 1-alpha-hydr
28	41.5	44.6	508	20	AAW89553 Human 25-hydroxyvl
29	41.5	44.6	508	22	AAW73908 Human 25-hydroxyvl
30	41	44.1	217	21	AAW40906 Human ORFX ORF670
31	41	44.1	490	19	AAW57899 Protein of clone C
32	41	44.1	490	21	AAW08441 Amino acid sequenc
33	41	44.1	490	22	AAW90716 Human CII480.9 pr
34	41	44.1	571	20	AAW36845 Protein involved i
35	40.5	43.5	17	16	AAW69772 Thrombospondin pep
36	40	43.0	17	16	AAW75811 Antimicrobial tach
37	40	43.0	17	16	AAW75813 Antimicrobial tach
38	40	43.0	17	16	AAW75814 Antimicrobial tach
39	40	43.0	20	19	AAW47705 Antimicrobial pept
40	40	43.0	20	20	AAW32638 Antimicrobial pept
41	40	43.0	27	19	AAW47694 Antimicrobial pept
42	40	43.0	27	20	AAW32628 Antimicrobial pept
43	40	43.0	79	22	AAW24132 Human EST encoded
44	40	43.0	172	22	AAW25307 Human protein sequ
45	40	43.0	272	21	AAW97289 Lipid associated p

ALIGNMENTS

RESULT 1  
AAW99417  
ID AAW99417 standard; peptide; 17 AA.  
AC AAW99417;  
DT 08-JUN-1999 (first entry)  
XX Tachyplesin derivative peptide SM2192.  
DE Linear: tachyplesin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.  
XX Synthetic.  
OS XX  
FH Key Location/Qualifiers  
FT Modified-site 17 /label= Nle  
XX WO9907728-A2.  
XX 18-FEB-1999.  
XX PD  
XX PF 06-AUG-1998; 98WO-FR01757.  
XX PR 12-AUG-1997; 97FR-0010297.  
XX PA (SYNT-) SYNT:EM SA.  
XX PI Calas B, Chavanieu A, Grassy G, Kaczorek M;  
XX DR WPI; 1999-190034/16.

Wed Feb 13 07:52:42 2002

CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 17 AA;

Query Match 66.7%; Score 62; DB 20; Length 17;  
 Best Local Similarity 75.0%; Pred. No. 0.0034;  
 Matches 12; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYAWRVHARGIRWLLR 16

Db 1 kvafrvayrgiryllr 16

RESULT 3

AAW47706  
 ID AAW47706 standard; peptide; 20 AA.

XX  
 AC AAW47706;

XX  
 DT 26-MAY-1998 (first entry)

XX  
 DE Antimicrobial peptide SUP1 analogue.

XX  
 KW Antimicrobial; transmembrane protein; TM; lentivirus lytic peptide; LLP;  
 KW SLP; amphipathic; antibacterial; antifungal; antiviral; antiprotozoal.

XX  
 OS Synthetic.

XX  
 OS Simian immunodeficiency virus.

XX  
 PN US5714577-A.

XX  
 PD 03-FEB-1998.

XX  
 PF 24-JAN-1997; 97US-0786748.

XX  
 PR 26-JAN-1996; 96US-0010634.

XX  
 PR 24-JAN-1997; 97US-0786748.

XX  
 PA (UYPI-) UNIV PITTSBURGH.

XX  
 PI Mietzner TA, Montelaro RC, Tencza SB;

XX  
 DR WPI; 1998-158352/14.

XX  
 PT Retroviral TM peptides - useful as antibacterial agents

XX  
 PS Disclosure; Column 10; 59pp; English.

XX  
 CC The invention relates to new antimicrobial peptides which correspond to  
 CC amino acid sequences in the transmembrane proteins of lentiviruses, in  
 CC particular HIV and SIV. These peptides comprise arginine rich sequences  
 CC which, when modelled for secondary structure, display high  
 CC amphipathicity and hydrophobic moment. Also disclosed are structural  
 CC and functional analogues and homologues of these peptides which also  
 CC display antimicrobial activity. The peptides are highly inhibitory to  
 CC microorganisms (bacteria, fungi, viruses and protozoa) but significantly  
 CC less toxic to red blood cells and other normal mammalian cells. Activity  
 CC is demonstrated against Gram positive and negative bacteria including  
 CC *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Enterococcus faecalis* and  
 CC *Serratia marcescens*.  
 CC The present sequence is one of 169 disclosed specific examples of  
 CC the new peptides. It is an analogue of the peptide designated SUP1  
 CC (see AAW47616) which is a peptide from the transmembrane protein of  
 CC SIV strain MM239.

XX Sequence 20 AA;

Query Match

48.4%; Score 45; DB 19; Length 20;

XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX Claim 8; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the tachyplesin family of  
 CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 17 AA;

Query Match 98.9%; Score 92; DB 20; Length 17;

Best Local Similarity 100.0%; Pred. No. 9.3e-08;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYAWRVHARGIRWLLR 16

Db 1 kvafrvayrgiryllr 16

RESULT 2

AAW99416  
 ID AAW99416 standard; peptide; 17 AA.

XX  
 AC AAW99416;

XX  
 DT 08-JUN-1999 (first entry)

XX  
 DE Tachyplesin derivative peptide SM2191.

XX  
 KW Linear; tachyplesin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX  
 OS Synthetic.

XX  
 PN WO9907728-A2.

XX  
 PD 18-FEB-1999.

XX  
 PF 06-AUG-1998; 98WO-FR01757.

XX  
 PR 12-AUG-1997; 97FR-0010297.

XX  
 PA (SYNT-) SYNT:EM SA.

XX  
 PI Calas B, Chavanleu A, Grassy G, Kaczorek M;

XX  
 DR WPI; 1999-190034/16.

XX  
 PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX Claim 8; Page 28; 37pp; French.

XX This peptide represents a linear derivative of the tachyplesin family of  
 CC peptide antibiotics. Tachyplesin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,

Best Local Similarity 58.3%; Pred. No. 1.5;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 WRVAHGRIRWLL 15  
ID AAY32548 standard; peptide; 20 AA.  
DB 2 wrllrrgrwil 13

RESULT 4  
AAY32548  
ID AAY32548 standard; peptide; 20 AA.  
XX AC AAY32548;

XX 21-OCT-1999 (first entry)

DE Antimicrobial peptide SLP-1 analogue.

KW Antimicrobial peptide; LLP1; SLP-1; LLP2; SLP2A; SLP2B; ELP; infection;  
KW growth inhibitor; microorganism; virus; gene therapy; vector production;  
KW sterilisation.

XX Synthetic.

XX US5945507-A.

XX 31-AUG-1999.

XX 18-SEP-1997; 97US-0932682.

XX 26-JAN-1996; 96US-0010634.

XX 24-JAN-1997; 97US-0786748.

XX 18-SEP-1997; 97US-0932682.

XX (UYPI-) UNIV PITTSBURGH.

XX Mletzner TA, Montelaro RC, Tencza SB;

XX WPI; 1999-508189/42.

XX Antimicrobial peptides useful for treating microbial infections

XX Claim 1; Column 107; 62pp; English.

CC This sequence represents an antimicrobial peptide of the invention, and  
CC is an analogue of the peptide SLP-1 (see AAY32551). The peptides can be  
CC used for treating infections caused by *Staphylococcus aureus*,  
CC methicillin resistant *S. aureus*, *Pseudomonas aeruginosa*, *Enterococcus*  
CC *faecalis*, *S. marcescens*, *Escherichia coli*, fungi, protozoa and viruses in  
CC a mammalian host. They can be used to inhibit growth of diverse  
CC microorganisms such as bacteria, fungi, protozoa and DNA and RNA viruses  
CC and can be used in tissue culture to inhibit unwanted microbial growth,  
CC particularly for the production of recombinant proteins or vectors for  
CC gene therapy. They can also be used in preventing infections through the  
CC sterilisation of wounds prior to suture and to sterilise surgical  
CC instruments. The unique structure of these antimicrobial peptides  
CC imparts high potency while selectivity is maintained, they are  
CC moderately haemolytic but only lyse red blood cells at high  
CC concentrations unlike melittin, a peptide extracted from bee venom, which  
CC is highly active against bacteria and lyses red blood cells showing  
CC little selectivity. The peptides target a membrane structure which makes  
CC it more difficult for a microorganism to develop a mechanism of  
CC resistance against this type of antibiotic. Their small size makes them  
CC relatively simple to prepare by standard synthetic peptide chemistry.

XX Sequence 20 AA;

Query Match 48.4%; Score 45; DB 20; Length 20;  
Best Local Similarity 58.3%; Pred. No. 1.5;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 WRVAHGRIRWLL 15

DB 2 wrllrrgrwil 13

RESULT 5  
AAW47695  
ID AAW47695 standard; peptide; 27 AA.  
XX AC AAW47695;

XX 26-MAY-1998 (first entry)

DE Antimicrobial peptide SLP1 analogue.

KW Antimicrobial; transmembrane protein; TM; lentivirus lytic peptide; LLP;  
KW SLP; amphoteric; antibacterial; antifungal; antiviral; antiprotozoal.

XX Synthetic.

XX Simian immunodeficiency virus.

XX US5714577-A.

XX 03-FEB-1998.

XX 24-JAN-1997; 97US-0786748.

XX 26-JAN-1996; 96US-0010634.

XX 24-JAN-1997; 97US-0786748.

XX (UYPI-) UNIV PITTSBURGH.

XX Mletzner TA, Montelaro RC, Tencza SB;

XX WPI; 1998-158352/14.

XX Retroviral TM peptides - useful as antibacterial agents

XX Disclosure; Column 10; 59pp; English.

CC The invention relates to new antimicrobial peptides which correspond to  
CC amino acid sequences in the transmembrane proteins of lentiviruses, in  
CC particular HIV and SIV. These peptides comprise arginine rich sequences  
CC which, when modelled for secondary structure, display high  
CC amphipathicity and hydrophobic moment. Also disclosed are structural  
CC and functional analogues and homologues of these peptides which also  
CC display antimicrobial activity. The peptides are highly inhibitory to  
CC microorganisms (bacteria, fungi, viruses and protozoa) but significantly  
CC less toxic to red blood cells and other normal mammalian cells. Activity  
CC is demonstrated against gram positive and negative bacteria including  
CC *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Enterococcus faecalis* and  
CC *Serratia marcescens*.  
CC The present sequence is one of 169 disclosed specific examples of  
CC the new peptides. It is an analogue of the peptide designated SLP1  
CC (see AAW47616) which is a peptide from the transmembrane protein of  
CC SIV strain WM239.

XX Sequence 27 AA;

Query Match 48.4%; Score 45; DB 19; Length 27;  
Best Local Similarity 58.3%; Pred. No. 2.1;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 WRVAHGRIRWLL 15

DB 2 wrllrrgrwil 13

RESULT 6

AAY32629

XX ID AAY32629 standard; peptide; 27 AA.

XX AAY32629;

XX 21-OCT-1999 (first entry)  
XX Antimicrobial peptide SLP-1 analogue.  
DE Antimicrobial peptide; LLP1; SLP-1; LLP2; SLP2A; SLP2B; ELP; Infection;  
XX growth inhibitor; microorganism; virus; gene therapy; vector production;  
KW sterilisation.  
KW  
XX  
XX Synthetic.  
OS Simian immunodeficiency virus.  
XX  
XX US5945507-A.  
XX  
XX 31-AUG-1999.  
PD  
XX  
XX 18-SEP-1997; 97US-0932682.  
PF  
XX  
XX 26-JAN-1996; 96US-0010634.  
PR  
XX 24-JAN-1997; 97US-0786748.  
PR  
XX 18-SEP-1997; 97US-0932682.  
PR  
XX (UYP1-) UNIV PITTSBURGH.  
XX  
XX Mietzner TA, Montelaro RC, Tencza SB;  
PI  
XX WPI; 1999-508189/42.  
XX  
XX Antimicrobial peptides useful for treating microbial infections  
PT  
XX  
XX Disclosure; Column 11; 62pp; English.  
PS  
XX  
XX This sequence represents an antimicrobial peptide of the invention, and  
CC is an analogue of the peptide SLP-1 (see AAY32551). The peptides can be  
CC used for treating infections caused by Staphylococcus aureus,  
CC methicillin resistant S. aureus, Pseudomonas aeruginosa, Enterococcus  
CC faecalis, S. marcescens, Escherichia coli, fungi, protozoa and viruses in  
CC a mammalian host. They can be used to inhibit growth of diverse  
CC microorganisms such as bacteria, fungi, protozoa and DNA and RNA viruses  
CC and can be used in tissue culture to inhibit unwanted microbial growth,  
CC particularly for the production of recombinant proteins or vectors for  
CC gene therapy. They can also be used in preventing infections through the  
CC sterilisation of wounds prior to suture and to sterilise surgical  
CC instruments. The unique structure of these antimicrobial peptides  
CC imparts high potency while selectivity is maintained, they are  
CC moderately haemolytic but only lyse red blood cells at high  
CC concentrations unlike melittin, a peptide extracted from bee venom, which  
CC is highly active against bacteria and lyses red blood cells showing  
CC little selectivity. The peptides target a membrane structure which makes  
CC it more difficult for a microorganism to develop a mechanism of  
CC resistance against this type of antibiotic. Their small size makes them  
CC relatively simple to prepare by standard synthetic peptide chemistry.  
XX  
XX Sequence 27 AA;  
SQ  
Query Match 48.4%; Score 45; DB 20; Length 27;  
Best Local Similarity 58.3%; Pred. NO. 2.1;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 4 WRVAHGRIRWLL 15  
Db 2 wrllrrgrwll 13  
RESULT 7  
AAM41298  
ID AAM41298 standard; Protein; 337 AA.  
XX  
XX AAM41298;  
AC  
XX  
XX 22-OCT-2001 (first entry)  
DT  
XX

DE Human polypeptide SEQ ID NO 6229.  
XX  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
OS  
XX WO200153312-A1.  
PN  
XX 26-JUL-2001.  
PD  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
PF  
XX  
XX 21-JAN-2000; 2000US-0488725.  
PR  
XX 25-APR-2000; 2000US-0552317.  
PR  
XX 09-JUL-2000; 2000US-0598042.  
PR  
XX 19-JUL-2000; 2000US-0620312.  
PR  
XX 03-AUG-2000; 2000US-0653450.  
PR  
XX 14-SEP-2000; 2000US-0662191.  
PR  
XX 19-OCT-2000; 2000US-0693036.  
PR  
XX 29-NOV-2000; 2000US-0727344.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
PI  
XX WPI; 2001-442253/47.  
DR  
XX N-ESDB; AAI60454.  
DR  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
PT  
XX Example 2; SEQ ID NO 6229; 10078pp; English.  
PS  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
XX the encoded polypeptides (AAM38642-AAAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX Sequence 337 AA;  
SQ  
Query Match 48.4%; Score 45; DB 22; Length 337;  
Best Local Similarity 41.7%; Pred. NO. 29;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 4 WRVAHGRIRWLL 15  
Db 250 wiiphkavrwl 261  
RESULT 8  
AAB44510  
ID AAB44510 standard; Protein; 202 AA.  
XX



AC AAB44510;  
 XX  
 DT 06-FEB-2001 (first entry)  
 XX  
 DE Plant viral movement protein SEQ ID 34.  
 XX  
 KW Plant viral movement protein; transport; transgenic plant;  
 KW viral resistance; cosuppression.  
 XX  
 OS Zea mays.  
 XX  
 PN W0200060088-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 06-APR-2000; 2000WO-US09110.  
 XX  
 PR 07-APR-1999; 99US-0128092.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Krebbers E, Weng Z, Cahoon RE;  
 XX  
 DR WPI; 2000-638467/61.  
 DR N-PSDB; AAC79364.  
 XX  
 PS Novel viral movement polypeptides and polynucleotides useful in field  
 XX of plant molecular biology, for producing transgenic plants, to prepare  
 XX antibodies and in immunological screening of cDNA expression libraries  
 XX  
 PS Claim 23; Page 52; 62pp; English.  
 XX  
 CC Polynucleotide sequences AAC79348-C79375 encode plant viral movement  
 CC proteins AAB44494-B44520. Some plant viruses have been shown to be able  
 CC to establish systemic infections via movement proteins that utilise  
 CC existing plant pathways to traffic macromolecules to surrounding cells.  
 CC Proteins such as those of the invention are similar to viral movement  
 CC proteins that function in the transport of nucleic acids from cell to  
 CC cell. The plant viral movement proteins are useful for obtaining a  
 CC nucleic acid fragment encoding a viral movement protein. Polynucleotides  
 CC encoding the plant viral movement proteins are useful for positive  
 CC selection of a transformed cell. The proteins are useful in the field of  
 CC plant molecular biology, and in the preparation of antibodies against the  
 CC proteins. The proteins are also useful for isolating cDNAs and genes  
 CC encoding homologous proteins from the same or other plant species, and to  
 CC create transgenic plants in which the protein is presented at higher or  
 CC lower levels than normal or in cell types or developmental stages in  
 CC which they are not normally found. The proteins and nucleotide sequences  
 CC may be used to control cosuppression and engineer plant virus  
 CC resistance.  
 XX  
 SQ Sequence 202 AA;

Query Match 47.8%; Score 44.5; DB 21; Length 202;  
 Best Local Similarity 47.6%; Pred. No. 21;  
 Matches 10; Conservative 3; Mismatches 3; Indels 5; Gaps 1;

QY 1 KYAWRVA-----HRGIRWLLR 16  
 ||||| :||| :|||  
 Db 35 kypwrvlglsrnrgrvffllr 55

RESULT 9  
 AAW47690  
 ID AAW47690 standard; peptide; 27 AA.  
 XX  
 AC AAW47690;  
 XX  
 DT 26-MAY-1998 (first entry)  
 XX  
 DE Antimicrobial peptide SLP1 analogue.

XX  
 KW Antimicrobial; transmembrane protein; TM; lentivirus lytic peptide; LLP;  
 KW SLP; amphoteric; antibacterial; antifungal; antiviral; antiprotozoal.  
 XX  
 OS Synthetic.  
 OS Simian immunodeficiency virus.  
 XX  
 PN US5714577-A.  
 XX  
 PD 03-FEB-1998.  
 XX  
 PF 24-JAN-1997; 97US-0786748.  
 XX  
 PR 26-JAN-1996; 96US-0010634.  
 PR 24-JAN-1997; 97US-0786748.  
 XX  
 PA (UYPI-) UNIV PITTSBURGH.  
 XX  
 PI Mietzner TA, Montelaro RC, Tencza SB;  
 XX WPI; 1998-158352/14.  
 DR  
 XX Retroviral TM peptides - useful as antibacterial agents  
 PT  
 XX Disclosure; Column 10; 59pp; English.  
 PS  
 XX The invention relates to new antimicrobial peptides which correspond to  
 CC amino acid sequences in the transmembrane proteins of lentiviruses, in  
 CC particular HIV and SIV. These peptides comprise arginine rich sequences  
 CC which, when modelled for secondary structure, display high  
 CC amphipathicity and hydrophobic moment. Also disclosed are structural  
 CC and functional analogues and homologues of these peptides which also  
 CC display antimicrobial activity. The peptides are highly inhibitory to  
 CC microorganisms (bacteria, fungi, viruses and protozoa) but significantly  
 CC less toxic to red blood cells and other normal mammalian cells. Activity  
 CC is demonstrated against Gram positive and negative bacteria including  
 CC *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Enterococcus faecalis* and  
 CC *Serratia marcescens*.  
 CC The present sequence is one of 169 disclosed specific examples of  
 CC the new peptides. It is an analogue of the peptide designated SLP1  
 CC (see AAW47616) which is a peptide from the transmembrane protein of  
 CC SIV strain MM239.  
 XX  
 SQ Sequence 27 AA;

Query Match 47.3%; Score 44; DB 19; Length 27;  
 Best Local Similarity 58.3%; Pred. No. 3;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 WRVAHRGIRWLL 15  
 ||||| ||| |||  
 Db 2 WtIrrgrwll 13

RESULT 10  
 AAY32624  
 ID AAY32624 standard; peptide; 27 AA.  
 XX  
 AC AAY32624;  
 XX  
 DT 21-OCT-1999 (first entry)  
 XX  
 DE Antimicrobial peptide SLP-1 analogue.  
 XX  
 KW Antimicrobial peptide; LLP1; SLP-1; LLP2; SLP2A; SLP2B; ELP; Infection;  
 KW growth inhibitor; microorganism; virus; gene therapy; vector production;  
 KW sterilisation.  
 XX  
 OS Synthetic.  
 OS Simian immunodeficiency virus.  
 XX  
 PN US5945507-A.

XX 31-AUG-1999.  
XX 18-SEP-1997; 97US-0932682.  
XX 26-JAN-1996; 96US-0010634.  
XX 24-JAN-1997; 97US-0786748.  
XX 18-SEP-1997; 97US-0932682.  
XX (UYPI-) UNIV PITTSBURGH.  
XX Mietzner TA, Montelaro RC, Tencza SB;  
XX WPI; 1999-508189/42.  
XX Antimicrobial peptides useful for treating microbial infections  
XX  
XX Disclosure: Column 11; 62pp; English.  
XX This sequence represents an antimicrobial peptide of the invention, and  
XX is an analogue of the peptide SLP-1 (see AAV32551). The peptides can be  
XX used for treating infections caused by *Staphylococcus aureus*,  
XX methicillin resistant *S. aureus*, *Pseudomonas aeruginosa*, *Enterococcus*  
XX *faecalis*, *S. marcescens*, *Escherichia coli*, fungi, protozoa and viruses in  
XX a mammalian host. They can be used to inhibit growth of diverse  
XX microorganisms such as bacteria, fungi, protozoa and DNA and RNA viruses  
XX and can be used in tissue culture to inhibit unwanted microbial growth,  
XX particularly for the production of recombinant proteins or vectors for  
XX gene therapy. They can also be used in preventing infections through the  
XX sterilisation of wounds prior to suture and to sterilise surgical  
XX instruments. The unique structure of these antimicrobial peptides  
XX imparts high potency while selectivity is maintained, they are  
XX moderately haemolytic but only lyse red blood cells at high  
XX concentrations unlike melittin, a peptide extracted from bee venom, which  
XX is highly active against bacteria and lyses red blood cells showing  
XX little selectivity. The peptides target a membrane structure which makes  
XX it more difficult for a microorganism to develop a mechanism of  
XX resistance against this type of antibiotic. Their small size makes them  
XX relatively simple to prepare by standard synthetic peptide chemistry.  
XX  
XX Sequence 27 AA;

Query Match 47.3%; Score 44; DB 20; Length 27;  
Best Local Similarity 58.3%; Pred. No. 3;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 WRVAHGRIRWLL 15  
II II II II II  
Db 2 wrtlrrggrwll 13

RESULT 11  
AAB95059  
ID AAB95059 standard; Protein; 125 AA.  
XX  
XX AAB95059;  
XX  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:16879.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX

PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 16879; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesising 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesising polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
XX Sequence 125 AA;

Query Match 47.3%; Score 44; DB 22; Length 125;  
Best Local Similarity 38.5%; Pred. No. 15;  
Matches 5; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 WRVAHGRIRWLLR 16  
I::I I I I I  
Db 37 wkichlgrfwirr 49

RESULT 12  
AAB94945  
ID AAB94945 standard; Protein; 298 AA.  
XX  
XX AAB94945;  
XX  
XX 21-AUG-2000 (first entry)  
XX A murine cholesterol 25-hydroxylase polypeptide.

DE Cholesterol 25-hydroxylase; serum cholesterol; immunogen.  
XX Mus sp.  
XX WO2000023596-A1.  
XX 27-APR-2000.  
XX 22-OCT-1999; 99WO-US24873.  
XX

XX 22-OCT-1998; 98US-0177419.  
 PR (TEXA ) UNIV TEXAS SYSTEM.  
 PA Russell DW, Lund EG;  
 XX WPI: 2000-339695/29.  
 DR N-PSDB; AAA15142.  
 XX New cholesterol 25-hydroxylase polypeptides, useful for regulating  
 PT cholesterol 25-hydroxylase activity and therefore serum cholesterol in  
 PT a mammalian host, have cholesterol 25-hydroxylase-specific structure  
 PT and activity -  
 XX  
 XX Claim 1; Page 47-48; 56pp; English.  
 PS The present sequence represents a murine cholesterol 25-hydroxylase  
 XX polypeptide. The polypeptides and polynucleotides can be used to  
 CC regulate cholesterol 25-hydroxylase activity and therefore serum  
 CC cholesterol in a mammalian host. The polypeptides, polynucleotides  
 CC and the cholesterol 25-hydroxylase-specific binding agents are useful  
 CC in diagnosis (e.g. genetic hybridization screens for cholesterol  
 CC 25-hydroxylase transcripts), therapy (e.g. cholesterol 25-hydroxylase  
 CC inhibitors to modulate serum cholesterol) and in the biopharmaceutical  
 CC industry (e.g. as immunogens, reagents for isolating natural  
 CC 25-hydroxylase genes and transcripts, reagents for screening chemical  
 CC libraries for lead pharmacological agents).  
 XX  
 SQ Sequence 298 AA;

Query Match 47.3%; Score 44; DB 21; Length 298;  
 Best Local Similarity 40.0%; Pred. No. 37;  
 Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 2 YAMRVAHRGIRLLR 16  
 : : : : :  
 Db 140 fawhlhkhkpwlyr 154

RESULT 13  
 AAR75806  
 ID AAR75806 standard; peptide; 17 AA.  
 AC AAR75806;  
 XX  
 XX 07-FEB-1996 (first entry)  
 DT Antimicrobial tachyplesin peptide derivative.  
 XX  
 XX Tachyplesin; inhibitor; antimicrobial; plant pathogen; fungi;  
 KW infection.  
 KW  
 XX Synthetic.  
 OS  
 XX WO9516776-A1.  
 PN  
 XX 22-JUN-1995.  
 PD  
 XX 19-DEC-1994; 94WO-US14619.  
 PF  
 XX 17-DEC-1993; 93US-0168809.  
 PR  
 XX (PION-) PIONEER HI-BRED INT INC.  
 PA Putman RJ, Rao AG, Rao A;  
 XX WPI: 1995-231570/30.  
 DR  
 XX New peptide derivs. of tachyplesin - having antimicrobial activity,  
 PT used against plant pathogenic fungi or human or animal infections  
 XX

PS Claim 1; Page 29; 45pp; English.  
 XX  
 CC AAR75802-04 and AAR75806-22 are peptide derivatives of tachyplesin  
 CC (AAR75805) a small peptide isolated from Japanese horseshoe crab  
 CC haemocytes which has antimicrobial properties. The peptide derivatives  
 CC also have antimicrobial activity and can be used for killing and  
 CC inhibiting fungi, in particular for fungi pathogenic to plants e.g.  
 CC Fusarium graminearum, Fusarium moniliforme, Sclerotinia sclerotiorum,  
 CC Sclerotinia trifoliorum and Aspergillus flavus. The peptides can  
 CC also be used for treating and preventing infection in humans and  
 CC animals.  
 XX  
 SQ Sequence 17 AA;

Query Match 46.2%; Score 43; DB 16; Length 17;  
 Best Local Similarity 61.5%; Pred. No. 2.6;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KYAWRVAHRGIRW 13  
 : : : : :  
 Db 1 kwatrvayrglay 13

RESULT 14  
 AAY69610  
 ID AAY69610 standard; peptide; 17 AA.  
 XX  
 AC AAY69610;  
 XX  
 XX 08-MAY-2000 (first entry)  
 DT Tachyplesin analogue, TPA.  
 XX  
 XX Tachyplesin analogue; generic; antimicrobial; disulphide bond;  
 KW antifungal; antiviral; antimicrobial; transgenic plant.  
 KW  
 XX Synthetic.  
 OS  
 XX  
 PH Key Location/Qualifiers  
 FT Modified-site 17 /note= "C-terminal amide"  
 FT  
 XX US6015941-A.  
 PN  
 XX 18-JAN-2000.  
 PD  
 XX 31-OCT-1997; 97US-0962034.  
 PF  
 XX 31-OCT-1997; 97US-0962034.  
 PR (PION-) PIONEER HI-BRED INT INC.  
 XX Rao AG;  
 XX  
 PI WPI: 2000-126327/11.  
 DR  
 XX New tachyplesin analogs useful for controlling fungal and bacterial  
 PT activity in agricultural and medical applications and for controlling  
 PT plant viruses have four cysteine substitutions -  
 PT  
 XX Example 1; Page -; 17pp; English.  
 PS  
 XX Sequences AAY69610-Y69611 and AAY69617 represent tachyplesin analogues  
 CC used in an exemplification of the present invention, in which the  
 CC native tachyplesin cysteine residues are replaced with Ala, Leu and  
 CC Asp, respectively. Tachyplesin (AAY69608) is a naturally occurring  
 CC antimicrobial peptide which contains two disulphide bonds which help  
 CC to maintain its tertiary structure. The invention relates to novel  
 CC peptide analogues of tachyplesin (Y696912-AAY69614, AAY69616) in which  
 CC the cysteine residues at positions 3, 7, 12 and 16 of the native  
 CC tachyplesin are replaced by the hydrophobic amino acids isoleucine,  
 CC valine, methionine, phenylalanine or tyrosine, the same amino acid being

CC present at all four positions. Despite being unable to form  
 CC intramolecular disulphide bonds, the analogues are functional as  
 CC antimicrobial agents. The tachyplesin analogues are useful for  
 CC controlling fungal and viral activity in agricultural and medical  
 CC applications and for controlling plant viruses. They can also be  
 CC expressed in transgenic plants, preferably wheat, sorghum, sunflower,  
 CC soya or especially maize plants to provide resistance to pathogenic fungi  
 CC and viruses. Note: The present sequence is not shown in the  
 CC specification, but is derived from the generic tachyplesin analogue  
 CC sequence given in column 23.

XX Sequence 17 AA;

Query Match 46.2%; Score 43; DB 21; Length 17;  
 Best Local Similarity 61.5%; Pred. No. 2.6;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 KYAWRVVHGRW 13  
 |:|:|:|:|:|:  
 Db 1 kwafrvaygiay 13

# RESULT 15

AAW47701  
 ID AAW47701 standard; peptide; 20 AA.

XX AC AAW47701;

XX DT 26-MAY-1998. (first entry)

XX DE Antimicrobial peptide SLP1 analogue.

XX KW Antimicrobial; transmembrane protein; TM; lentivirus lytic peptide; LLP;  
 KW SLP; amphoteric; antibacterial; antifungal; antiviral; antiprotozoal.

XX OS Synthetic.

XX OS Simian immunodeficiency virus.

XX PN US5714577-A.

XX PD 03-FEB-1998.

XX PF 24-JAN-1997; 97US-0786748.

XX PR 26-JAN-1996; 96US-0010634.

XX PR 24-JAN-1997; 97US-0786748.

XX PA (UYPI-) UNIV PITTSBURGH.

PI Mietzner TA, Montelaro RC, Tencza SB;

XX WPI; 1998-158352/14.

XX Retroviral TM peptides - useful as antibacterial agents

PS Disclosure; Column 10; 59pp; English.

XX The invention relates to new antimicrobial peptides which correspond to  
 CC amino acid sequences in the transmembrane proteins of lentiviruses, in  
 CC particular HIV and SIV. These peptides comprise arginine rich sequences  
 CC which, when modelled for secondary structure, display high  
 CC amphipathicity and hydrophobic moment. Also disclosed are structural  
 CC and functional analogues and homologues of these peptides which also  
 CC display antimicrobial activity. The peptides are highly inhibitory to  
 CC microorganisms (bacteria, fungi, viruses and protozoa) but significantly  
 CC less toxic to red blood cells and other normal mammalian cells. Activity  
 CC is demonstrated against Gram positive and negative bacteria including  
 CC *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Enterococcus faecalis* and  
 CC *Serratia marcescens*.  
 CC The present sequence is one of 169 disclosed specific examples of  
 CC the new peptides. It is an analogue of the peptide designated SLP1  
 CC (see AAW47616) which is a peptide from the transmembrane protein of

CC SIV strain MM239.  
 XX Sequence 20 AA;

Query Match 45.2%; Score 42; DB 19; Length 20;  
 Best Local Similarity 50.0%; Pred No. 4.4;  
 Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 4 WRVAHGRWLL 15  
 | | | | |  
 Db 2 wetlrrgrwil 13

Search completed: February 12, 2002, 12:30:34  
 Job time: 367 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:25 ; Search time 106.12 seconds  
(without alignments)  
3.605 Million cell updates/sec

Title: US-09-485-571-33

Perfect score: 93

Sequence: 1 KYAWRVHRGIRLLRX 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	48.4	20	1	US-08-786-748A-97
2	45	48.4	20	2	US-08-932-682-97
3	45	48.4	27	1	US-08-786-748A-86
4	45	48.4	27	2	US-08-932-682-86
5	44	47.3	27	1	US-08-786-748A-81
6	44	47.3	27	2	US-08-932-682-81
7	44	47.3	481	4	US-08-943-714-10
8	43	46.2	17	1	US-08-168-809-5
9	42	45.2	20	1	US-08-786-748A-92
10	42	45.2	20	1	US-08-786-748A-94
11	42	45.2	20	2	US-08-932-682-92
12	42	45.2	20	2	US-08-932-682-94
13	42	45.2	21	1	US-08-786-748A-89
14	42	45.2	21	2	US-08-932-682-89
15	42	45.2	27	1	US-08-786-748A-80
16	42	45.2	27	2	US-08-932-682-80
17	42	45.2	28	1	US-08-786-748A-76
18	42	45.2	28	2	US-08-932-682-76
19	41.5	44.6	508	4	US-09-111-730-2
20	40	43.0	17	1	US-08-168-809-10
21	40	43.0	17	1	US-08-168-809-12
22	40	43.0	17	1	US-08-168-809-13
23	40	43.0	20	1	US-08-786-748A-96
24	40	43.0	20	2	US-08-932-682-96
25	40	43.0	27	1	US-08-786-748A-85
26	40	43.0	27	2	US-08-932-682-85
27	40	43.0	729	1	US-08-971-937-2

Sequence 2, Appl  
Sequence 73, Appl  
Sequence 90, Appl  
Sequence 93, Appl  
Sequence 99, Appl  
Sequence 101, Appl  
Sequence 70, Appl  
Sequence 93, Appl  
Sequence 99, Appl  
Sequence 101, Appl  
Sequence 88, Appl  
Sequence 88, Appl  
Sequence 78, Appl  
Sequence 79, Appl  
Sequence 82, Appl  
Sequence 83, Appl  
Sequence 84, Appl

ALIGNMENTS

RESULT 1  
US-08-786-748A-97  
; Sequence 97, Application US/08786748A  
; Patent No. 5714577  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/786,748A  
; FILING DATE: 24-JAN-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/010,634  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-408-2500  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 97:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5714577e.  
; US-08-786-748A-97

Query Match 48.4%; Score 45; DB 1; Length 20;  
Best Local Similarity 58.3%; Pred. No. 0.66;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WRVAHGRWLL 15  
||: || ||:  
Db 2 WLLRRGGRWIL 13

## RESULT 2

US-08-932-682-97  
; Sequence 97, Application US/08932682  
; Patent No. 5945507

## GENERAL INFORMATION:

; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169

## CORRESPONDENCE ADDRESS:

; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10112-0228

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0

## CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/786,748

; FILING DATE: 24-JAN-1997

## ATTORNEY/AGENT INFORMATION:

; NAME: Rochelle K. Seide

; REGISTRATION NUMBER: 32,300

; REFERENCE/DOCKET NUMBER: AP30421-A

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-705-5000

; TELEFAX: 212-765-2519

## INFORMATION FOR SEQ ID NO: 97:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 20 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: No. 5945507e

US-08-932-682-97

Query Match 48.4%; Score 45; DB 2; Length 20;

Best Local Similarity 58.3%; Pred. No. 0.66;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WRVAHGRWLL 15  
||: || ||:  
Db 2 WLLRRGGRWIL 13

## RESULT 3

US-08-786-748A-86

; Sequence 86, Application US/08786748A

; Patent No. 5714577

## GENERAL INFORMATION:

; APPLICANT: Ronald, Montelaro C.

; APPLICANT: Tencza, Sarah B.

; APPLICANT: Mietzner, Timothy A.

; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES

; NUMBER OF SEQUENCES: 169

## CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10112-0228

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: FastSeq Version 2.0

; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0

## CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/786,748A

; FILING DATE: 24-JAN-1997

; CLASSIFICATION: 514

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/010,634

; FILING DATE: 26-JAN-1996

## ATTORNEY/AGENT INFORMATION:

; NAME: Rochelle K. Seide

; REGISTRATION NUMBER: 32,300

; REFERENCE/DOCKET NUMBER: AP30421

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-408-2500

; TELEFAX: 212-765-2519

; INFORMATION FOR SEQ ID NO: 86:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 27 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: No. 5714577e

US-08-786-748A-86

Query Match 48.4%; Score 45; DB 1; Length 27;

Best Local Similarity 58.3%; Pred. No. 0.89;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WRVAHGRWLL 15  
||: || ||:  
Db 2 WLLRRGGRWIL 13

## RESULT 4

US-08-932-682-86

; Sequence 86, Application US/08932682

; Patent No. 5945507

## GENERAL INFORMATION:

; APPLICANT: Ronald, Montelaro C.

; APPLICANT: Tencza, Sarah B.

; APPLICANT: Mietzner, Timothy A.

; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES

; NUMBER OF SEQUENCES: 169

## CORRESPONDENCE ADDRESS:

; ADDRESSEE: BAKER & BOTTS, L.L.P.

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10112-0228

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: FastSeq Version 2.0

## CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/932,682

; FILING DATE: 18-SEP-1997

; CLASSIFICATION: 530

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/786,748

; FILING DATE: 24-JAN-1997

## ATTORNEY/AGENT INFORMATION:

; NAME: Rochelle K. Seide



REGISTRATION NUMBER: 32,300  
REFERENCE/DOCKET NUMBER: AP30421-A  
TELEPHONE: 212-705-5000  
TELEFAX: 212-765-2519  
INFORMATION FOR SEQ ID NO: 86:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5945507e  
US-08-932-682-86

Query Match 48.4%; Score 45; DB 2; Length 27;  
Best Local Similarity 58.3%; Pred. No. 0.89;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WRVAHGRWLL 15  
II: II II: I  
Db 2 WRLRRGGRWIL 13

RESULT 5  
US-08-786-748A-81  
; Sequence 81, Application US/08786748A  
; Patent No. 5714577  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/786,748A  
; FILING DATE: 24-JAN-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/010,634  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421  
; TELEPHONE: 212-408-2500  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 81:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5714577e  
US-08-786-748A-81

Query Match 47.3%; Score 44; DB 1; Length 27;  
Best Local Similarity 58.3%; Pred. No. 1.3;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 WRVAHGRWLL 15  
II: II II: I  
Db 2 WRLRRGGRWIL 13

RESULT 6  
US-08-932-682-81  
; Sequence 81, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
; FILING DATE: 24-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-705-5000  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 81:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5945507e  
US-08-932-682-81

Query Match 47.3%; Score 44; DB 2; Length 27;  
Best Local Similarity 58.3%; Pred. No. 1.3;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 WRVAHGRWLL 15  
II: II II: I  
Db 2 WRLRRGGRWIL 13

RESULT 7  
US-08-943-714-10  
; Sequence 10, Application US/08943714  
; Patent No. 6187578  
; GENERAL INFORMATION:  
; APPLICANT: Blinkovsky, Alexander  
; APPLICANT: Berka, Randy  
; APPLICANT: Rey, Michael  
; APPLICANT: Golightly, Elizabeth  
; APPLICANT: Klotz, Alan  
; APPLICANT: Mathisen, Thomas Erik  
; APPLICANT: Dambmann, Claus

```

; REGISTRATION NUMBER: 29,342
; REFERENCE/DOCKET NUMBER: 0173R US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-245-3595
; TELEFAX: 515-245-3634
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-168-809-5

Query Match          46.2%; Score 43; DB 1; Length 17;
Best Local Similarity 61.5%; Pred. No. 1.1;
Matches      8; Conservative    4; Mismatches   1; Indels     0; Gaps     0;

QY      1 KYAWRVVAHGRIRW 13
        |.:|:|:|:|:|:|:|:|:
Db       1 KWAFRVAYRGIAI 13

RESULT      9
US-08-786-748A-92
; Sequence 92, Application US/08786748A
; Patent No. 5714577
; GENERAL INFORMATION:
; APPLICANT: Ronald, Montelaro C.
; APPLICANT: Tencza, Sarah B.
; APPLICANT: Mietzner, Timothy A.
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,748A
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/010,634
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Rochelle K. Seide
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: AP30421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2500
; TELEFAX: 212-765-2519
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 5714577e
US-08-786-748A-92

Query Match          45.2%; Score 42; DB 1; Length 20;

```

Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 10  
US-08-786-748A-94  
; Sequence 94, Application US/08786748A  
; Patent No. 5714577  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/786,748A  
; FILING DATE: 24-JAN-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/010,634  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421  
; TELEPHONE: 212-408-2500  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5714577e  
US-08-786-748A-94

Query Match 45.2%; Score 42; DB 1; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 11  
US-08-932-682-92  
; Sequence 92, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
; FILING DATE: 24-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421-A  
; TELEPHONE: 212-705-5000  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5945507e  
US-08-932-682-92

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 12  
US-08-932-682-94  
; Sequence 94, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
US-08-932-682-94

Query Match 45.2%; Score 42; DB 1; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 13  
US-08-932-682-94  
; Sequence 94, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
US-08-932-682-94

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 14  
US-08-932-682-94  
; Sequence 94, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
US-08-932-682-94

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
; FILING DATE: 24-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421-A  
; TELEPHONE: 212-705-5000  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5945507e  
US-08-932-682-92

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 15  
US-08-932-682-94  
; Sequence 94, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
; FILING DATE: 24-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421-A  
; TELEPHONE: 212-705-5000  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 92:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5945507e  
US-08-932-682-92

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 16  
US-08-932-682-94  
; Sequence 94, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
US-08-932-682-94

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 17  
US-08-932-682-94  
; Sequence 94, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
US-08-932-682-94

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
Db 2 WETLRGCRWIL 13

RESULT 18  
US-08-932-682-94  
; Sequence 94, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
US-08-932-682-94

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

; FILING DATE: 24-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-705-5000  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5945507e  
US-08-932-682-94

Query Match 45.2%; Score 42; DB 2; Length 20;  
Best Local Similarity 50.0%; Pred. No. 1.9; Indels 5; Mismatches 1; Conservative 6; Gaps 0;

QY 4 WRVAHGRIRWLL 15  
| | | | |  
Db 2 WETLRGCRWIL 13

RESULT 13  
US-08-786-748A-89  
; Sequence 89, Application US/08786748A  
; Patent No. 5714577  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/786,748A  
; FILING DATE: 24-JAN-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/010,634  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-408-2500  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 89:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5714577e  
US-08-786-748A-89

Query Match 45.2%; Score 42; DB 1; Length 21;  
Best Local Similarity 50.0%; Pred. No. 2;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 4 WRVAHGRIRWLL 15  
| | | | |  
Db 3 WETLRGCRWIL 14

RESULT 14  
US-08-932-682-89  
; Sequence 89, Application US/08932682  
; Patent No. 5945507  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES  
; NUMBER OF SEQUENCES: 169  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BAKER & BOTTS, L.L.P.  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,682  
; FILING DATE: 18-SEP-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/786,748  
; FILING DATE: 24-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rochelle K. Seide  
; REGISTRATION NUMBER: 32,300  
; REFERENCE/DOCKET NUMBER: AP30421-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-705-5000  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 89:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5945507e  
US-08-932-682-89

Query Match 45.2%; Score 42; DB 2; Length 21;  
Best Local Similarity 50.0%; Pred. No. 2;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 4 WRVAHGRIRWLL 15  
| | | | |  
Db 3 WETLRGCRWIL 14

RESULT 15  
US-08-786-748A-80  
; Sequence 80, Application US/08786748A  
; Patent No. 5714577  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Montelaro C.  
; APPLICANT: Tencza, Sarah B.  
; APPLICANT: Mietzner, Timothy A.  
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES

NUMBER OF SEQUENCES: 169  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112-0228  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/786,748A  
FILING DATE: 24-JAN-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/010,634  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Rochelle K. Seide  
REGISTRATION NUMBER: 32,300  
REFERENCE/DOCKET NUMBER: AP30421  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-408-2500  
TELEFAX: 212-765-2519  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5714577e  
US-08-786-748A-80

Query Match 45.2%; Score 42; DB 1; Length 27;  
Best Local Similarity 50.0%; Pred. No. 2.5;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 WRVAHGRWLL 15  
| | | | |  
Db 2 WETLRGCRWIL 13

Search completed: February 12, 2002, 12:32:25  
Job time: 453 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:37 : Search time 126.85 Seconds  
(without alignments)  
10.809 Million cell updates/sec

Title: US-09-485-571-20  
Perfect score: 33  
Sequence: 1 XGGXXXXXXXXXXXXXG 18  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	48.5	342	B33061	homeobox protein p
2	16	48.5	347	A54863	homeobox protein -
3	16	48.5	361	E86692	N-acetylmuramyl-L
4	16	48.5	605	T33913	hypothetical prote
5	16	48.5	742	C34734	transcription fact
6	16	48.5	825	B34734	sterol regulatory
7	16	48.5	1139	B54962	acidic ribosomal p
8	15	45.5	110	R5BYA1	19K antigen - Myco
9	15	45.5	163	S22630	chorion protein s3
10	15	45.5	306	S08607	beta-lactamase (EC
11	15	45.5	311	JN0520	beta-1,3-glucanase
12	15	45.5	377	T50563	pre-B-cell leukemi
13	15	45.5	430	A56002	PBX2 protein - mou
14	15	45.5	430	T09061	phosphatidylinosit
15	15	45.5	490	T09084	paired box transcr
16	15	45.5	520	T78502	CAD ATPase (AAAI),
17	15	45.5	523	B96835	protein kinase sgg
18	15	45.5	575	S35327	hypothetical prote
19	15	45.5	1067	S35423	gene hind sight pro
20	15	45.5	1075	T48805	odz protein - frui
21	15	45.5	1541	T02831	tenascin-like prot
22	15	45.5	1772	A45532	neuroglian, long c
23	15	45.5	1891	T13594	hypothetical prote
24	15	45.5	1920	T13893	gene
25	15	45.5	2406	A54148	hypothetical prote
26	15	45.5	2515	S47008	hypothetical prote
27	14	42.4	120	PH0083	hypothetical prote
28	14	42.4	205	S55670	hypothetical prote
29	14	42.4	236	C69060	hypothetical prote

30	14	42.4	270	2	S65739	basigin precursor
31	14	42.4	367	2	G85362	hypothetical prote
32	14	42.4	376	2	A48840	extracellular prote
33	14	42.4	419	2	D83465	conserved hypothet
34	14	42.4	427	2	A32372	female-specific do
35	14	42.4	448	2	T35667	ammonium transport
36	14	42.4	463	2	T36810	probable integral
37	14	42.4	493	2	E71725	glutamyl-tRNA amid
38	14	42.4	537	2	B33485	spore coat protein
39	14	42.4	549	2	B32372	male-specific doub
40	14	42.4	556	2	T42100	serine/threonine p
41	14	42.4	556	2	T36502	serine/threonine p
42	14	42.4	571	1	UMFF	period clock prote
43	14	42.4	600	2	S07638	spore coat protein
44	14	42.4	647	2	T39141	hypothetical prote
45	14	42.4	737	2	S47857	basic protein, cyt

ALIGNMENTS

RESULT 1  
B33061  
homeotic protein prl - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Jan-1990 #sequence\_revision 17-Jan-1990 #text\_change 17-Oct-1997  
C:Accession: B33061  
R:Nourse, J.; Mellenin, J.D.; Galili, N.; Wilkinson, J.; Stanbridge, E.; Smith, S.  
submitted to the Protein Sequence Database, January 1990  
A:Reference number: A33061  
A:Accession: B33061  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-342 <NOU>  
C:Genetics:  
A:Gene: GDB:PBX1  
A:Cross-references: GDB:125351; OMIM:176310  
A:Map position: 1q23-1q23  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: acute lymphoblastic leukemia; DNA binding; homeobox; nucleus; transcript  
F:146-205/Domain: homeobox homology #status atypical <HOX>

Query Match 48.5%; Score 16; DB 2; Length 342;  
Best Local Similarity 17.6%; Pred. No. 5.6e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXXXXXXG 18  
||  
Db 36 GGSAAAAAASGGAG 52

RESULT 2  
A54863  
homeobox protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 24-Sep-1999  
C:Accession: A54863  
R:Kagawa, N.; Ogo, A.; Takahashi, Y.; Iwamatsu, A.; Waterman, M.R.  
J. Biol. Chem. 269, 18716-18719, 1994  
A:Title: A cAMP-regulatory sequence (CRS1) of CYP17 is a cellular target for the h  
A:Reference number: A54863; MUID:94308119  
A:Accession: A54863  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-347 <RES>  
A:Cross-references: GB:I27453; NID:g456108; PIDN:AAA21832.1; PID:g456109  
C:Genetics:  
A:Gene: PBX1B  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:234-293/Domain: homeobox homology <HOX>

```

Query Match      48.5%; Score 16; DB 2; Length 347;
Best Local Similarity 17.6%; Pred. No. 5.7e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY  2 GGXXXXXXXG 18
    ||
Db   124 GGSAAAAAAGGAG 140

RESULT 3
E86692
N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) [imported] - Lactococcus lactis subsp.
N; Alternate names: N-acetylmuramidase
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: E86692
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissensbach, J.; Ehrlich
  Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: E86692
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <STO>
A:Cross-references: GB:AE005176; NID:g12723426; PIDN:AAK04639.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: acmD
C:Keywords: hydrolase

Query Match      48.5%; Score 16; DB 2; Length 361;
Best Local Similarity 17.6%; Pred. No. 5.9e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY  2 GGXXXXXXXG 18
    ||
Db   177 GGTAVTTSSASTNSAG 193

RESULT 4
T33913
Hypothetical protein Y25C1A.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33913
R:Kalicki, J.; Smith, A.; Gibson, A.
  Submitted to the EMBL Data Library, February 1999
A:Description: The sequence of C. elegans cosmid Y25C1A.
A:Reference number: Z21437
A:Accession: T33913
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-605 <KAL>
A:Cross-references: EMBL:AF125459; PIDN:AAI2833.1; GSPDB:GN00020; CESP:Y25C1A.3
A:Experimental source: strain Bristol N2; clone Y25C1A
C:Genetics:
A:Gene: CESP:Y25C1A.3
A:Map position: 2
A:Introns: 392/3; 429/1; 567/3

Query Match      48.5%; Score 16; DB 2; Length 605;
Best Local Similarity 17.6%; Pred. No. 8.8e+03;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY  2 GGXXXXXXXG 18
    ||
Db   346 GGSTATTAAAGGSTASTG 362

RESULT 5

```

```

C34734
transcription factor 3/homeotic protein prl mutant fusion protein II - human
N:Alternate names: E2A/prl mutant fusion protein II; E2A-Prl mutant fusion protein c
C:Species: Homo sapiens (man)
C:Date: 13-Jul-1990 #sequence_revision 17-Nov-1995 #text_change 22-May-1997
C:Accession: C34734; B34733
R:Kamps, M.P.; Murre, C.; Sun, X.; Baltimore, D.
  Cell 60, 547-555, 1990
A:Title: A new homeobox gene contributes the DNA binding domain of the t(1;19) trans
  location mutation.
A:Reference number: A34734; MUID:90150282
A:Accession: C34734
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-742 <KAM>
A:Cross-references: GB:M31522; NID:g339894
R:Nourse, J.; Mellentin, J.D.; Galili, N.; Wilkinson, J.; Stanbridge, E.; Smith, S.D.
  Cell 60, 535-545, 1990
A:Title: Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion
  protein.
A:Reference number: A34733; MUID:90150281
A:Accession: B34733
A:Molecule type: mRNA
A:Residues: 7-579, 'DE', 582-742 <NOU>
A:Cross-references: GB:M31222; NID:g181905
C:Comment: This sequence is the chimeric product of a translocation mutation.
C:Genetics:
A:Gene: TCF3/PBX1
C:Keywords: fusion protein

Query Match      48.5%; Score 16; DB 4; Length 742;
Best Local Similarity 17.6%; Pred. No. 1e+04;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY  2 GGXXXXXXXG 18
    ||
Db   519 GGSAAAAAAGGAG 535

RESULT 6
B34734
transcription factor 3/homeotic protein prl mutant fusion protein I - human
N:Alternate names: E2A/prl mutant fusion protein I
C:Species: Homo sapiens (man)
C:Date: 13-Jul-1990 #sequence_revision 17-Nov-1995 #text_change 20-Apr-2000
C:Accession: B34734; A34733
R:Kamps, M.P.; Murre, C.; Sun, X.; Baltimore, D.
  Cell 60, 547-555, 1990
A:Title: A new homeobox gene contributes the DNA binding domain of the t(1;19) trans
  location mutation.
A:Reference number: A34734; MUID:90150282
A:Accession: B34734
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-825 <KAM>
A:Cross-references: GB:M31522; NID:g339894; PIDN:AAA36764.1; PID:g339895
R:Nourse, J.; Mellentin, J.D.; Galili, N.; Wilkinson, J.; Stanbridge, E.; Smith, S.D.
  Cell 60, 535-545, 1990
A:Title: Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion
  protein.
A:Reference number: A34733; MUID:90150281
A:Accession: A34733
A:Molecule type: mRNA
A:Residues: 7-579, 'DE', 582-825 <NOU>
A:Cross-references: GB:M31170; NID:gl90357; PIDN:AAA36484.1; PID:gl90358
C:Comment: This sequence is the chimeric product of a translocation mutation.
C:Genetics:
A:Gene: TCF3/PBX1
C:Keywords: fusion protein

Query Match      48.5%; Score 16; DB 4; Length 825;
Best Local Similarity 17.6%; Pred. No. 1.1e+04;
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY  2 GGXXXXXXXG 18

```



Db 519 GGSAAAAAAGGAG 535  
||  
RESULT 7  
B54962  
sterol regulatory element binding protein 2 precursor - Chinese hamster  
C:Species: Cricetulus griseus (Chinese hamster)  
C:Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 04-Sep-1998  
C:Accession: B54962  
R:Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S.  
Genes Dev. 8, 1910-1919, 1994  
A:Title: Sterol-resistant transcription in CHO cells caused by gene rearrangement that t  
A:Reference number: A54962; MUID:95047343  
A:Accession: B54962  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1139 <YAN>  
A:Cross-references: GB:U12330  
A:Note: 493 Ser was also found  
C:Superfamily: sterol regulatory element binding protein  
C:Keywords: DNA binding; membrane protein

Query Match 48.5%; Score 16; DB 2; Length 1139;  
Best Local Similarity 17.6%; Pred. No. 1.4e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 GGXXXXXXXG 18  
||  
Db 57 GGSXXXXSSSSSSG 73

RESULT 8  
R5BVA1  
acidic ribosomal protein P2.e.B, cytosolic - yeast (Saccharomyces cerevisiae)  
N:Alternate names: acidic ribosomal protein P2.alpha; protein D9481.1; protein YDR382w;  
C:Species: Saccharomyces cerevisiae  
C:Date: 31-Oct-1980 #sequence\_revision 31-Mar-1992 #text\_change 21-Jul-2000  
C:Accession: A35109; A28104; A02776; S61177; S69666  
R:Newton, C.H.; Shimmis, L.C.; Yee, J.; Dennis, P.P.  
J. Bacteriol. 172, 579-588, 1990  
A:Title: A family of genes encode the multiple forms of the Saccharomyces cerevisiae rib  
1 protein.  
A:Reference number: A35109; MUID:90130289  
A:Accession: A35109  
A:Molecule type: DNA  
A:Residues: 1-110 <NEW>  
A:Cross-references: GB:M26505; NID:g171810; PIDN:AAA34732.1; PID:g171811  
R:Remacha, M.; Saenz-Robles, M.T.; Vilella, M.D.; Ballesta, J.P.G.  
J. Biol. Chem. 263, 9094-9101, 1988  
A:Title: Independent genes coding for three acidic proteins of the large ribosomal subun  
A:Reference number: A92726; MUID:88243786  
A:Accession: A28104  
A:Molecule type: DNA  
A:Residues: 1-110 <REM>  
A:Cross-references: EMBL:J03761; NID:g172397; PIDN:AAA34972.1; PID:g172398  
A:Note: the authors translated the codon GAA for residue 28 as Ala  
R:Itoh, T.  
Biochim. Biophys. Acta 671, 16-24, 1981  
A:Title: Primary structure of an acidic ribosomal protein YPA1 from Saccharomyces cerevi  
A:Reference number: A02776; MUID:82069169  
A:Accession: A02776  
A:Molecule type: protein  
A:Residues: 1-74, 'GPAS', 79-85, 'G', 86-90, 92-110 <ITO>  
R:Ding, H.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of S. cerevisiae cosmid 9481.  
A:Reference number: S61159  
A:Accession: S61177  
A:Molecule type: DNA  
A:Residues: 1-110 <DIN>  
A:Cross-references: EMBL:U28373; NID:g849184; PIDN:AAB64818.1; PID:g849203; GSPDB:GN0000

A:Experimental source: strain S288C (AB972)  
R:Dietrich, F.S.  
submitted to the EMBL Data Library, July 1995  
A:Description: The sequence of S. cerevisiae cosmids 9481, 9509, 9926, 9461, and 1a  
A:Reference number: S69665  
A:Accession: S69666  
A:Molecule type: DNA  
A:Residues: 1-110 <DIE>  
A:Cross-references: EMBL:U32274; NID:g927313; PIDN:AAB64824.1; PID:g927315; GSPDB:C  
C:Genetics:  
A:Gene: SGD:RPL45; RPLA4; MIPS:YDR382w  
A:Cross-references: SGD:S0002790; MIPS:YDR382w  
A:Map position: 4R  
C:Superfamily: rat acidic ribosomal protein p1  
C:Keywords: phosphoprotein; protein biosynthesis; ribosome  
F:1-110/Product: acidic ribosomal protein P2.e.B #status experimental <MAT>

Query Match 45.5%; Score 15; DB 1; Length 110;  
Best Local Similarity 17.6%; Pred. No. 3.9e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 GGXXXXXXXG 18  
||  
Db 69 GCASSAAAGAAAGG 85

RESULT 9  
S22630  
19K antigen - Mycobacterium intracellulare  
N:Alternate names: gene MI22 protein  
C:Species: Mycobacterium intracellulare  
C:Date: 03-May-1994 #sequence\_revision 20-Feb-1995 #text\_change 22-Oct-1999  
C:Accession: S22630  
R:Nair, J.; Rouse, D.A.; Morris, S.L.  
Mol. Microbiol. 6, 1431-1439, 1992  
A:Title: Nucleotide sequence analysis and serologic characterization of the Mycobac  
A:Reference number: S22630; MUID:92326626  
A:Accession: S22630  
A:Molecule type: DNA  
A:Residues: 1-163 <NAI>  
A:Cross-references: EMBL:X65483; NID:g44361; PIDN:CAA46469.1; PID:g581330  
A:Note: the authors translated the initiation codon GTG for residue 1 as Val  
C:Genetics:  
A:Gene: MI22  
A:Start codon: GTG

Query Match 45.5%; Score 15; DB 2; Length 163;  
Best Local Similarity 17.6%; Pred. No. 5.3e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 GGXXXXXXXG 18  
||  
Db 24 GGNKSGTSASSSSG 40

RESULT 10  
S08607  
chorion protein 38 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 02-Dec-1993 #sequence\_revision 01-Dec-1995 #text\_change 21-Jul-2000  
C:Accession: S08607  
R:Spradling, A.C.; de Cicco, D.V.; Wakimoto, B.T.; Levine, J.F.; Kafayan, L.J.; C  
EMBO J. 6, 1045-1053, 1987  
A:Title: Amplification of the x-linked Drosophila chorion gene cluster requires a  
A:Reference number: S07193; MUID:87246506  
A:Accession: S08607  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-306 <SPR>  
A:Cross-references: EMBL:X05245; NID:g7725; PIDN:CAA28871.1; PID:g7727  
A:Note: the authors translated the codon TAC for residue 112 as Tyr

## C:Genetics:

A:Gene: FlyBase:Cp38  
A:Cross-references: FlyBase:FBgn0000360  
A:Introns: 15/3

Query Match 45.5%; Score 15; DB 2; Length 306;  
Best Local Similarity 17.6%; Pred. No. 8.7e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 39 CGADAASAAAAAGGAG 55

## RESULT 11

JN0520  
beta-lactamase (EC 3.5.2.6) - Streptomyces cellulosa  
C:Species: Streptomyces cellulosa  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 20-Jun-2000  
C:Accession: JN0520  
R:Ogawara, H.  
Gene 124, 111-114, 1993  
A:Title: Sequence of a gene encoding beta-lactamase form Streptomyces cellulosa.  
A:Reference number: JN0520; MUID:93178958  
A:Accession: JN0520  
A:Molecule type: DNA  
A:Residues: 1-311 <OGA>  
A:Cross-references: GB:D12653; NID:g287453; PIDN:BAA02176.1; PID:g287454  
C:Superfamily: beta-lactamase I  
C:Keywords: hydrolase  
F:86.184/Active site: Ser, Glu #status predicted

Query Match 45.5%; Score 15; DB 2; Length 311;  
Best Local Similarity 17.6%; Pred. No. 8.8e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 21 GGALALGTTTASASAG 37

## RESULT 12

T50563  
beta-1,3-glucanase (EC 3.2.1.-) precursor [imported] - rice  
C:Species: Oryza sativa (rice)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 28-Jul-2000  
C:Accession: T50563  
R:Romero, G.O.; Doan, M.; Rodriguez, R.L.  
submitted to the EMBL Data Library, September 1996  
A:Description: Rice beta-1,3-glucanase, Gns9.  
A:Reference number: 225133  
A:Accession: T50563  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-377 <ROM>  
A:Cross-references: EMBL:U72255; PIDN:AAD10386.1  
A:Experimental source: strain M202  
C:Genetics:

A:Gene: Gns9  
A:Introns: 25/1  
C:Superfamily: beta-1,3-glucanase  
C:Keywords: glycosidase; hydrolase

Query Match 45.5%; Score 15; DB 2; Length 377;  
Best Local Similarity 17.6%; Pred. No. 1e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 353 GGNLAAAAAARTTAG 369

## RESULT 13

A56002  
pre-B-cell leukemia transcription factor 2 - human  
N:Alternate names: homeotic protein PBX2  
C:Species: Homo sapiens (man)  
C:Date: 12-Apr-1995 #sequence\_revision 12-Apr-1995 #text\_change 24-Sep-1999  
C:Accession: A56002; S19009; A55562  
R:Aguado, B.; Campbell, R.D.  
Genomics 25, 650-659, 1995  
A:Title: The novel gene GL7, located in the human major histocompatibility complex,  
A:Reference number: A56002; MUID:95278934  
A:Accession: A56002  
A:Molecule type: DNA  
A:Residues: 1-430 <AGU>  
A:Cross-references: GB:X80700; NID:g625185; PIDN:CAA56717.1; PID:g634053  
R:Monica, K.; Galili, N.; Nourse, J.; Saltman, D.; Cleary, M.L.  
Mol. Cell. Biol. 11, 6149-6157, 1991  
A:Title: PBX2 and PBX3, new homeobox genes with extensive homology to the human prot  
A:Reference number: S19009; MUID:92049345  
A:Accession: S19009  
A:Molecule type: mRNA  
A:Residues: 1-392, 'I', 394-430 <MON>  
A:Cross-references: EMBL:X59842; NID:g35312; PIDN:CAA42503.1; PID:g35313  
R:Sugaya, K.; Fukagawa, T.; Matsumoto, K.; Mita, K.; Takahashi, E.; Ando, A.; Inoko,  
Genomics 23, 408-419, 1994  
A:Title: Three genes in the human MHC class III region near the junction with the cl  
nterpart of mouse mammary tumor gene int-3.  
A:Reference number: A55562; MUID:95137587  
A:Accession: A55562  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-430 <RES>  
A:Cross-references: GB:D28769; NID:g561657; PIDN:BAA05957.1; PID:g561658  
C:Genetics:

A:Gene: GDB:PBX2; G17  
A:Cross-references: GDB:306356; OMIM:176311  
A:Map position: 6p21.3-6p21.3  
A:Introns: 74/2; 99/1; 181/3; 245/2; 290/3; 342/1; 371/3; 400/3  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:245-304/Domain: homeobox homology <HOX>

Query Match 45.5%; Score 15; DB 2; Length 430;  
Best Local Similarity 17.6%; Pred. No. 1.1e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 133 GGSAAAAAASGGG 149

## RESULT 14

T09061  
PBX2 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jan-2000  
C:Accession: T09061  
R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.  
submitted to the EMBL Data Library, October 1997  
A:Description: Sequence of the mouse major histocompatibility locus class III region  
A:Reference number: Z16543  
A:Accession: T09061  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-430 <ROW>  
A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564949  
C:Genetics:  
A:Gene: PBX2  
A:Map position: 17  
A:Introns: 74/2; 99/1; 181/3; 245/2; 290/3; 342/1; 371/3; 400/3

C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:245-304/Domain: homeobox homology <HOX>

Query Match 45.5%; Score 15; DB 2; Length 430;  
Best Local Similarity 17.6%; Pred. No. 1.1e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 133 GGGSAAGAGAGG 149

RESULT 15  
T09084  
phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)  
C:Species: Chlamydomonas reinhardtii  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T09084  
R:Molendijk, A.J.; Irvine, R.F.  
Plant Mol. Biol. 37, 53-66, 1998  
A:Title: Inositide signalling in Chlamydomonas: Characterization of a phosphatidylinositol  
A:Reference number: Z16411; MUID:98281574  
A:Accession: T09084  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-490 <MOL>  
A:Cross-references: EMBL:U97663; NID:g2109290; PIDN:AAC50018.1; PID:g2109291  
A:Experimental source: strain cw-15  
C:Genetics:  
A:Introns: 265/3; 331/3; 370/3; 455/1; 481/3

Query Match 45.5%; Score 15; DB 2; Length 490;  
Best Local Similarity 17.6%; Pred. No. 1.3e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 422 GGGAAAGAGAGG 438

Search completed: February 12, 2002, 12:34:38  
Job time: 556 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:22 ; Search time 106.12 Seconds  
(without alignments)  
3.817 Million cell updates/sec

Title: US-09-485-571-20  
Perfect score: 33  
Sequence: 1 XGXXXXXXXXXXXXXG 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/ptodata/2/laa/5A\_COMB.pep:\*  
2: /cgn2.6/ptodata/2/laa/5B\_COMB.pep:\*  
3: /cgn2.6/ptodata/2/laa/6A\_COMB.pep:\*  
4: /cgn2.6/ptodata/2/laa/6B\_COMB.pep:\*  
5: /cgn2.6/ptodata/2/laa/PCTUS\_COMB.pep:\*  
6: /cgn2.6/ptodata/2/laa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	51.5	182	1	US-08-240-712-29
2	17	51.5	182	1	US-08-443-890-29
3	17	51.5	738	3	US-08-864-038A-3
4	16	48.5	18	1	US-08-240-712-33
5	16	48.5	18	1	US-08-443-890-33
6	16	48.5	801	1	US-07-906-349A-6
7	15	45.5	349	4	US-09-105-390-48
8	15	45.5	377	4	US-09-105-390-64
9	14	42.4	204	1	US-08-185-432-8
10	14	42.4	303	1	US-08-185-432-5
11	14	42.4	341	2	US-08-209-521-11
12	14	42.4	362	2	US-08-985-090-5
13	14	42.4	362	3	US-09-165-543-32
14	14	42.4	445	3	US-09-165-543-5
15	14	42.4	606	4	US-09-247-806-6
16	14	42.4	737	1	US-08-185-432-2
17	14	42.4	737	1	US-08-185-432-4
18	14	42.4	1127	4	US-09-150-460B-11
19	13	39.4	31	1	US-08-425-069-46
20	13	39.4	31	2	US-08-317-844B-46
21	13	39.4	31	4	US-08-556-978B-26
22	13	39.4	31	4	US-08-556-978B-69
23	13	39.4	31	4	US-08-556-978B-89
24	13	39.4	34	1	US-08-425-069-42
25	13	39.4	34	1	US-08-425-069-54
26	13	39.4	34	2	US-08-317-844B-42
27	13	39.4	34	2	US-08-317-844B-54
					Sequence 29, Appl
					Sequence 3, Appl
					Sequence 33, Appl
					Sequence 6, Appl
					Sequence 48, Appl
					Sequence 64, Appl
					Sequence 8, Appl
					Sequence 5, Appl
					Sequence 11, Appl
					Sequence 5, Appl
					Sequence 32, Appl
					Sequence 5, Appl
					Sequence 6, Appl
					Sequence 2, Appl
					Sequence 4, Appl
					Sequence 11, Appl
					Sequence 46, Appl
					Sequence 11, Appl
					Sequence 46, Appl
					Sequence 26, Appl
					Sequence 69, Appl
					Sequence 89, Appl
					Sequence 42, Appl
					Sequence 54, Appl
					Sequence 42, Appl
					Sequence 54, Appl

28	13	39.4	101	4	US-08-556-978B-20	Sequence 20, Appl
29	13	39.4	101	4	US-08-556-978B-22	Sequence 22, Appl
30	13	39.4	101	4	US-08-556-978B-62	Sequence 62, Appl
31	13	39.4	101	4	US-09-247-806-3	Sequence 3, Appl
32	13	39.4	101	4	US-09-247-806-5	Sequence 5, Appl
33	13	39.4	101	4	US-09-247-806-7	Sequence 7, Appl
34	13	39.4	118	3	US-08-301-162-10	Sequence 10, Appl
35	13	39.4	129	4	US-09-135-994-12	Sequence 12, Appl
36	13	39.4	251	1	US-08-209-747-8	Sequence 8, Appl
37	13	39.4	251	1	US-08-458-298-8	Sequence 8, Appl
38	13	39.4	338	1	US-08-218-686-2	Sequence 2, Appl
39	13	39.4	338	3	US-08-460-242-2	Sequence 2, Appl
40	13	39.4	401	6	5252556-1	Patent No. 5252556
41	13	39.4	445	2	US-08-985-090-2	Sequence 2, Appl
42	13	39.4	445	3	US-09-165-543-2	Sequence 2, Appl
43	13	39.4	445	4	US-09-167-354-7	Sequence 7, Appl
44	13	39.4	490	1	US-08-672-571A-1	Sequence 1, Appl
45	13	39.4	604	4	US-08-556-978B-63	Sequence 63, Appl

ALIGNMENTS

RESULT 1  
US-08-240-712-29  
; Sequence 29, Application US/08240712  
; Patent No. 5599907  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, DAVID C.  
; APPLICANT: MATHEWS, ANTONY JAMES  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
; TITLE OF INVENTION: HEMOGLOBINS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,712  
FILING DATE: 09-MAY-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON=6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 182 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: /note= one or both of Gly  
OTHER INFORMATION: residues 131 and 132 can be absent; one or both  
OTHER INFORMATION: of Gly residues 147 and 148 can be absent  
US-08-240-712-29

Query Match 51.5%; Score 17; DB 1; Length 182;  
Best Local Similarity 17.6%; Pred. No. 1.4e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 130 GGGAAAAAAG 146

RESULT 2  
US-08-443-890-29  
; Sequence 29, Application US/08443890  
; Patent No. 5739011  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, DAVID C.  
; APPLICANT: MATHEWS, ANTONY JAMES  
; APPLICANT: STETLER, GARY L.  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
; HEMOGLOBINS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/443,890  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/240,712  
; FILING DATE: 09-MAY-1994  
; APPLICATION NUMBER: PCT/US92/09752  
; FILING DATE: 13-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: ANDERSON=6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633

INFORMATION FOR SEQ ID NO: 29:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 182 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:

OTHER INFORMATION: /note= one or both of Gly  
; OTHER INFORMATION: residues 131 and 132 can be absent: one or both  
; OTHER INFORMATION: of Gly residues 147 and 148 can be absent  
US-08-443-890-29

Query Match 51.5%; Score 17; DB 1; Length 182;  
Best Local Similarity 17.6%; Pred. No. 1.4e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 130 GGGAAAAAAG 146

RESULT 3  
US-08-864-038A-3  
; Sequence 3, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBO  
; TITLE OF INVENTION: TO SAID POLYPEPTIDE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 812-5 Hirano

; STREET: Isshinden  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: P-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)986-2340  
; TELEFAX: (212)953-7733  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 738

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Pinctada fucata  
; CELL TYPE: mantle epithelial cell  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: from 1 to 738  
; IDENTIFICATION METHOD: E (by experiment)

US-08-864-038A-3

Query Match 51.5%; Score 17; DB 3; Length 738;  
Best Local Similarity 17.6%; Pred. No. 3.5e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 280 GGSAAAAAAGG 296

RESULT 4  
US-08-240-712-33  
; Sequence 33, Application US/08240712  
; Patent No. 5599907  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, DAVID C.  
; APPLICANT: MATHEWS, ANTONY JAMES  
; APPLICANT: STETLER, GARY L.  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
; HEMOGLOBINS

NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,712  
FILING DATE: 09-MAY-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON-6  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: /note= one or both of Gly  
OTHER INFORMATION: residues 2 and 3 can be absent; one or both  
OTHER INFORMATION: of Gly residues 17 and 18 can be absent  
US-08-240-712-33

Query Match 48.5%; Score 16; DB 1; Length 18;  
Best Local Similarity 17.6%; Pred. No. 4.9e+02;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 1 GGGAAAAAAAAAAGG 17

RESULT 5  
US-08-443-890-33  
Sequence 33, Application US/08443890  
Patent No. 5739011  
GENERAL INFORMATION:  
APPLICANT: ANDERSON, DAVID C.  
APPLICANT: MATHEWS, ANTONY JAMES  
APPLICANT: STETLER, GARY L.  
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/443,890  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/240,712  
FILING DATE: 09-MAY-1994  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON-6  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: /note= one or both of Gly  
OTHER INFORMATION: residues 2 and 3 can be absent; one or both  
OTHER INFORMATION: of Gly residues 17 and 18 can be absent  
US-08-443-890-33

Query Match 48.5%; Score 16; DB 1; Length 18;  
Best Local Similarity 17.6%; Pred. No. 4.9e+02;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 1 GGGAAAAAAAAAAGG 17

RESULT 6  
US-07-906-349A-6  
Sequence 6, Application US/07906349A  
Patent No. 5434064  
GENERAL INFORMATION:  
APPLICANT: Schlessinger, Joseph  
APPLICANT: Skolnik, Edward Y.  
APPLICANT: Margolis, Benjamin L.  
TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR  
IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/906,349A  
FILING DATE: 30-JUN-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/643,237  
FILING DATE: 18-JAN-1991  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 801 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-906-349A-6

Query Match 48.5%; Score 16; DB 1; Length 801;  
Best Local Similarity 17.6%; Pred. No. 6.3e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 2 GGXXXXXXXG 18  
Db 457 GGTGGATTAAATG 473

RESULT 7  
US-09-105-390-48  
; Sequence 48, Application US/09105390  
; Patent No. 6288303  
; GENERAL INFORMATION:  
; APPLICANT: Rodriguez, Raymond  
; TITLE OF INVENTION: Rice Beta-Glucanase Enzymes  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Ave., Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/105,390  
; FILING DATE: Filed herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/050,675  
; FILING DATE: 25-JUN-97  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Petithory, Joanne R.  
; REGISTRATION NUMBER: P42,995  
; REFERENCE/DOCKET NUMBER: 2000-0455.30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-0880  
; TELEFAX: 650-324-0960  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 349 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-09-105-390-48

Query Match 45.5%; Score 15; DB 4; Length 349;  
Best Local Similarity 17.6%; Pred. No. 6.2e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 2 GGXXXXXXXG 18

Db 325 GGNLAAAAAARTTAG 341

RESULT 8  
US-09-105-390-64  
; Sequence 64, Application US/09105390  
; Patent No. 6288303  
; GENERAL INFORMATION:  
; APPLICANT: Rodriguez, Raymond  
; TITLE OF INVENTION: Rice Beta-Glucanase Enzymes  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Ave., Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/105,390  
; FILING DATE: Filed herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/050,675  
; FILING DATE: 25-JUN-97  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Petithory, Joanne R.  
; REGISTRATION NUMBER: P42,995  
; REFERENCE/DOCKET NUMBER: 2000-0455.30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-0880  
; TELEFAX: 650-324-0960  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 377 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-09-105-390-64

Query Match 45.5%; Score 15; DB 4; Length 377;  
Best Local Similarity 17.6%; Pred. No. 6.5e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 2 GGXXXXXXXG 18  
Db 353 GGNLAAAAAARTTAG 369

RESULT 9  
US-08-185-432-8  
; Sequence 8, Application US/08185432  
; Patent No. 5750652  
; GENERAL INFORMATION:  
; APPLICANT: Attavanis-Tsakonas, Spyridon  
; APPLICANT: Busseau, Isabelle  
; APPLICANT: Diederich, Robert J.  
; APPLICANT: Xu, Tian  
; APPLICANT: Matsuno, Kenji  
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
; ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 23



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: PENNIE & EDMONDS  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/185,432  
;; FILING DATE: 21-JAN-1994  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Misrock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 7326-006  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-8864/9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 204 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; US-08-185-432-8

Query Match 42.4% Score 14; DB 1; Length 204;  
Best Local Similarity 17.6% Pred. No. 7.5e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 GGXXXXXXXG 18  
||  
Db 18 GGSAASSCATMALSTAG 34

RESULT 10  
US-08-185-432-5  
; Sequence 5, Application US/08185432  
; Patent No. 5750652  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Busseau, Isabelle  
; APPLICANT: Diederich, Robert J.  
; APPLICANT: Xu, Tian  
; APPLICANT: Matsuno, Kenji  
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/185,432  
; FILING DATE: 21-JAN-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie

;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 7326-006  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-8864/9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 303 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; US-08-185-432-5

Query Match 42.4% Score 14; DB 1; Length 303;  
Best Local Similarity 17.6% Pred. No. 9.7e+03;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 GGXXXXXXXG 18  
||  
Db 18 GGSAASSCATMALSTAG 34

RESULT 11  
US-08-209-521-11  
; Sequence 11, Application US/08209521  
; Patent No. 5922855  
; GENERAL INFORMATION:  
; APPLICANT: Liskay, Robert M.  
; APPLICANT: Bronner, C. Eric  
; APPLICANT: Baker, Sean M.  
; APPLICANT: Bollag, Roni J.  
; APPLICANT: Kolodner, Richard D.  
; TITLE OF INVENTION: MAMMALIAN DNA MISMATCH REPAIR GENES  
; TITLE OF INVENTION: HMLH1 AND HPMS1  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &  
; ADDRESS: Heuser  
; STREET: 520 S.W. Yamhill, Suite 200  
; CITY: Portland  
; STATE: Oregon  
; COUNTRY: US  
; ZIP: 97204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/209,521  
; FILING DATE: 08-MAR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Van Rysseberghe, Pierre C.  
; REGISTRATION NUMBER: 33,557  
; REFERENCE/DOCKET NUMBER: OHSU 306A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (503) 224-6655  
; TELEFAX: (503) 295-6679  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 341 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; POSITION IN GENOME:  
; MAP POSITION: 3p21.3-23  
; US-08-209-521-11

Query Match 42.4%; Score 14; DB 2; Length 341;  
Best Local Similarity 17.6%; Pred. No. 1.e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 17 GGAAATGAGTAACATG 33

RESULT 12  
US-08-985-090-5  
; Sequence 5, Application US/08985090  
; Patent No. 5885893  
; GENERAL INFORMATION:  
; APPLICANT: Andrew D.J. Goodearl  
; TITLE OF INVENTION: MUSCARINIC RECEPTORS AND USES THEREFOR  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/985.090  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jean M. Silveri  
; REGISTRATION NUMBER: 39,030  
; REFERENCE/DOCKET NUMBER: MNI-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 362 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-985-090-5

Query Match 42.4%; Score 14; DB 2; Length 362;  
Best Local Similarity 17.6%; Pred. No. 1.e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 214 GGAAASPTSSGSSSRG 230

RESULT 13  
US-09-165-543-32  
; Sequence 32, Application US/09165543  
; Patent No. 6093545  
; GENERAL INFORMATION:  
; APPLICANT: Andrew D.J. Goodearl and Sandra Glucksman  
; TITLE OF INVENTION: Muscarinic Receptors and Uses Therefor  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts

; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/165,543  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/042,780  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elizabeth A. Hanley  
; REGISTRATION NUMBER: 33,505  
; REFERENCE/DOCKET NUMBER: MNI-032CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 362 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-165-543-32

Query Match 42.4%; Score 14; DB 3; Length 362;  
Best Local Similarity 17.6%; Pred. No. 1.e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 GGXXXXXXXG 18  
||  
Db 214 GGAAASPTSSGSSSRG 230

RESULT 14  
US-09-165-543-5  
; Sequence 5, Application US/09165543  
; Patent No. 6093545  
; GENERAL INFORMATION:  
; APPLICANT: Andrew D.J. Goodearl and Sandra Glucksman  
; TITLE OF INVENTION: Muscarinic Receptors and Uses Therefor  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/165,543  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/042,780  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elizabeth A. Hanley  
; REGISTRATION NUMBER: 33,505  
; REFERENCE/DOCKET NUMBER: MNI-032CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214

; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 445 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-165-543-5

Query Match 42.4%; Score 14; DB 3; Length 445;  
Best Local Similarity 17.6%; Pred. No. 1.2e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 2 GGXXXXXXXXXXXXXG 18  
||  
Db 297 GGAASPTSSGSSRG 313

RESULT 15  
US-09-247-806-6  
; Sequence 6, Application US/09247806  
; Patent No. 6280747  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; TITLE OF INVENTION: ANALOG  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/247,806  
; CURRENT FILING DATE: 1999-02-11  
; EARLIER APPLICATION NUMBER: FR 98/01614  
; EARLIER FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 606  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-247-806-6

Query Match 42.4%; Score 14; DB 4; Length 606;  
Best Local Similarity 17.6%; Pred. No. 1.5e+04;  
Matches 3; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 2 GGXXXXXXXXXXXXXG 18  
||  
Db 326 GGCGAAAAAAAGGAG 342

Search completed: February 12, 2002, 12:32:23  
Job time: 451 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:33 ; Search time 242.57 Seconds  
(without alignments)  
5.191 Million cell updates/sec

Title: US-09-485-571-31

Perfect score: 22

Sequence: 1 xxxxxxxxxxxxxxxxxxxx 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	27.3	2	22 AAB91687	Oploid peptide SEQ
2	6	27.3	2	22 AAB91688	Oploid peptide SEQ
3	6	27.3	2	22 AAB91721	Oploid peptide SEQ
4	6	27.3	2	22 AAB91731	Oploid peptide SEQ
5	6	27.3	2	22 AAB91740	Oploid peptide SEQ
6	6	27.3	2	22 AAB92149	Polypeptide SEQ ID
7	6	27.3	2	22 AAB92356	Miscellaneous pept
8	6	27.3	3	4 AAB30453	Sequence of collag
9	6	27.3	3	9 AAB81081	Sequence encoded b
10	6	27.3	3	10 AAB91660	Synthetic peptide
11	6	27.3	3	11 AAR04607	Antiviral agent.

12	6	27.3	3	11 AAR00718	Core repeat of cel
13	6	27.3	3	13 AAR23219	Biocompatible poly
14	6	27.3	3	14 AAR36707	Adhesion formation
15	6	27.3	3	14 AAR53144	RGD peptide deriva
16	6	27.3	3	15 AAR65950	Tripeptide inhibit
17	6	27.3	3	15 AAR44666	Platelet aggregati
18	6	27.3	3	15 AAR53243	Heparin derivative
19	6	27.3	3	15 AAR53243	Heparin derivative
20	6	27.3	3	15 AAR47430	ACE inhibiting pep
21	6	27.3	3	15 AAR51441	IGF-1 analogue N-t
22	6	27.3	3	15 AAR61092	ACE-inhibiting tri
23	6	27.3	3	15 AAR61095	ACE-inhibiting tri
24	6	27.3	3	16 AAR70472	Cancer metastasis
25	6	27.3	3	16 AAR71717	Potential cross-11
26	6	27.3	3	16 AAR64689	HPF3 peptide deriv
27	6	27.3	3	16 AAR62425	Accelerator peptid
28	6	27.3	3	17 AAR11094	Platelet-targeting
29	6	27.3	3	17 AAR98166	Peptide fragment o
30	6	27.3	3	17 AAR89785	Cell adhesion inh
31	6	27.3	3	18 AAR31143	Platelet-targeting
32	6	27.3	3	18 AAR25173	RGD-peptide capabl
33	6	27.3	3	18 AAR28457	PECAM-1 cyclic inh
34	6	27.3	3	18 AAR28437	PECAM-1 inhibitor
35	6	27.3	3	19 AAR48589	Integrin receptor
36	6	27.3	3	19 AAR56241	Anti-inflammatory
37	6	27.3	3	19 AAR56226	Anti-inflammatory
38	6	27.3	3	19 AAR56187	Anti-inflammatory
39	6	27.3	3	19 AAR56198	Anti-inflammatory
40	6	27.3	3	19 AAR52445	Loop region used i
41	6	27.3	3	19 AAR52447	Loop region used i
42	6	27.3	3	20 AAR43493	Linker for dual av
43	6	27.3	3	20 AAR43494	Linker for dual av
44	6	27.3	3	20 AAR39827	Elastin peptide fr
45	6	27.3	3	22 AAR92359	Miscellaneous pept

ALIGNMENTS

RESULT 1  
AAB91687  
ID AAB91687 standard; Peptide; 2 AA.

AC AAB91687;

DT 22-JUN-2001 (first entry)

DE Oploid peptide SEQ ID NO:863.

XX Protection; endogenous therapeutic peptide; peptidase; conjugation;  
KW blood component; modification; succinimidyl; maleimido group; amino;  
KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.

OS Homo sapiens.

OS Synthetic.

OS XX

PN WO200063900-A2.

XX XX

PD 23-NOV-2000.

XX 17-MAY-2000; 2000WO-US13576.

XX 17-MAY-1999; 99US-0134406.

PR 10-SEP-1999; 99US-0153406.

PR 15-OCT-1999; 99US-0159783.

XX (CONJ-) CONJUCHEM INC.

XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibadeau K;

PI WPI; 2001-112059/12.

XX Modifying and attaching therapeutic peptides to albumin prevents

PT

PT peptide degradation, useful for increasing length of in vivo activity  
 PT Disclosure; Page 477; 733pp; English.  
 XX  
 CC The present invention describes a modified therapeutic peptide (I)  
 CC comprising a therapeutically active amino acid region (III) and a  
 CC reactive group (II) (e.g. succinimide and maleimido groups) attached to  
 CC a less therapeutically active amino acid region (IV), which covalently  
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a  
 CC peptide stabilised therapeutic peptide composed of 3-50 amino acids.  
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth  
 CC factors and neurotransmitters, to protect them from peptidase activity  
 CC in vivo for the treatment of various disorders. Endogenous therapeutic  
 CC peptides are not suitable as drug candidates as they require frequent  
 CC administration due to rapid degradation by peptidases in the body.  
 CC Modifying and attaching therapeutic peptides to albumin prevents or  
 CC reduces the action of peptidases to increase length of activity (half  
 CC life) and specificity as bonding to large molecules decreases  
 CC intracellular uptake and interference with physiological processes.  
 CC AAB90829 to AAB92441 represent peptides which can be used in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 2 AA;

Query Match 27.3%; Score 6; DB 22; Length 2;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 G 10  
 Db 1 g 1

RESULT 2  
 AAB91688  
 ID AAB91688 standard; Peptide; 2 AA.  
 AC AAB91688;  
 XX  
 DT 22-JUN-2001 (first entry)  
 DE Opioid peptide SEQ ID NO:864.  
 KW Protection; endogenous therapeutic peptide; peptidase; conjugation;  
 KW blood component; modification; succinimide; maleimido group; amino;  
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200069900-A2.  
 PN 23-NOV-2000.  
 XX  
 PD 17-MAY-2000; 2000WO-US13576.  
 PF 17-MAY-1999; 99US-0134406.  
 PR 10-SEP-1999; 99US-0153406.  
 PR 15-OCT-1999; 99US-0159783.  
 XX (CONJ-) CONJUCHEM INC.  
 PA Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;  
 PI WPI; 2001-112059/12.  
 XX  
 DR Modifying and attaching therapeutic peptides to albumin prevents  
 XX peptidase degradation, useful for increasing length of in vivo activity  
 PT  
 PT Disclosure; Page 477; 733pp; English.  
 XX

XX  
 CC The present invention describes a modified therapeutic peptide (I)  
 CC comprising a therapeutically active amino acid region (III) and a  
 CC reactive group (II) (e.g. succinimide and maleimido groups) attached to  
 CC a less therapeutically active amino acid region (IV), which covalently  
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a  
 CC peptide stabilised therapeutic peptide composed of 3-50 amino acids.  
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth  
 CC factors and neurotransmitters, to protect them from peptidase activity  
 CC in vivo for the treatment of various disorders. Endogenous therapeutic  
 CC peptides are not suitable as drug candidates as they require frequent  
 CC administration due to rapid degradation by peptidases in the body.  
 CC Modifying and attaching therapeutic peptides to albumin prevents or  
 CC reduces the action of peptidases to increase length of activity (half  
 CC life) and specificity as bonding to large molecules decreases  
 CC intracellular uptake and interference with physiological processes.  
 CC AAB90829 to AAB92441 represent peptides which can be used in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 2 AA;

Query Match 27.3%; Score 6; DB 22; Length 2;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 G 10  
 Db 1 g 1

RESULT 3  
 AAB91721  
 ID AAB91721 standard; Peptide; 2 AA.  
 AC AAB91721;  
 XX  
 DT 22-JUN-2001 (first entry)  
 DE Opioid peptide SEQ ID NO:897.  
 KW Protection; endogenous therapeutic peptide; peptidase; conjugation;  
 KW blood component; modification; succinimide; maleimido group; amino;  
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200069900-A2.  
 PN 23-NOV-2000.  
 XX  
 PD 17-MAY-2000; 2000WO-US13576.  
 PF 17-MAY-1999; 99US-0134406.  
 PR 10-SEP-1999; 99US-0153406.  
 PR 15-OCT-1999; 99US-0159783.  
 XX (CONJ-) CONJUCHEM INC.  
 PA Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;  
 PI WPI; 2001-112059/12.  
 XX  
 DR Modifying and attaching therapeutic peptides to albumin prevents  
 XX peptidase degradation, useful for increasing length of in vivo activity  
 PT  
 PT Disclosure; Page 487; 733pp; English.  
 XX  
 CC The present invention describes a modified therapeutic peptide (I)  
 CC comprising a therapeutically active amino acid region (III) and a  
 CC reactive group (II) (e.g. succinimide and maleimido groups) attached to

CC a less therapeutically active amino acid region (IV), which covalently  
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a  
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids  
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth  
 CC factors and neurotransmitters, to protect them from peptidase activity  
 CC in vivo for the treatment of various disorders. Endogenous therapeutic  
 CC peptides are not suitable as drug candidates as they require frequent  
 CC administration due to rapid degradation by peptidases in the body.  
 CC Modifying and attaching therapeutic peptides to albumin prevents or  
 CC reduces the action of peptidases to increase length of activity (half  
 CC life) and specificity as bonding to large molecules decreases  
 CC intracellular uptake and interference with physiological processes.  
 CC AAB90829 to AAB92441 represent peptides which can be used in the  
 CC exemplification of the present invention.

XX Sequence 2 AA;

Query Match 27.3%; Score 6; DB 22; Length 2;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
 Db 2 g 2

RESULT 4

AAB91731  
 ID AAB91731 standard; Peptide; 2 AA.

XX AC AAB91731;

XX DT 22-JUN-2001 (first entry)

XX DE Opioid peptide SEQ ID NO:907.

XX Protection; endogenous therapeutic peptide; peptidase; conjugation;  
 KW blood component; modification; succinimidyl; maleimido group; amino;  
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.

XX Homo sapiens.  
 OS Synthetic.

XX WO200069900-A2.

XX PD 23-NOV-2000.

XX PF 17-MAY-2000; 2000WO-US13576.

XX PR 17-MAY-1999; 99US-0134406.

XX PR 10-SEP-1999; 99US-0153406.

XX PR 15-OCT-1999; 99US-0159783.

XX PA (CONJ-) CONJUCHEM INC.

XX PI Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;

XX WPI; 2001-112059/12.

PT Modifying and attaching therapeutic peptides to albumin prevents  
 PT peptidase degradation, useful for increasing length of in vivo activity

XX Disclosure; Page 490; 733pp; English.

XX The present invention describes a modified therapeutic peptide (I)  
 CC comprising a therapeutically active amino acid region (III) and a  
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to  
 CC a less therapeutically active amino acid region (IV), which covalently  
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a  
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.  
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth

CC factors and neurotransmitters, to protect them from peptidase activity  
 CC in vivo for the treatment of various disorders. Endogenous therapeutic  
 CC peptides are not suitable as drug candidates as they require frequent  
 CC administration due to rapid degradation by peptidases in the body.  
 CC Modifying and attaching therapeutic peptides to albumin prevents or  
 CC reduces the action of peptidases to increase length of activity (half  
 CC life) and specificity as bonding to large molecules decreases  
 CC intracellular uptake and interference with physiological processes.  
 CC AAB90829 to AAB92441 represent peptides which can be used in the  
 CC exemplification of the present invention.

XX Sequence 2 AA;

Query Match 27.3%; Score 6; DB 22; Length 2;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10  
 Db 2 g 2

RESULT 5

AAB91740

ID AAB91740 standard; Peptide; 2 AA.

XX AC AAB91740;

XX DT 22-JUN-2001 (first entry)

XX DE Opioid peptide SEQ ID NO:916.

XX Protection; endogenous therapeutic peptide; peptidase; conjugation;  
 KW blood component; modification; succinimidyl; maleimido group; amino;  
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.

XX Homo sapiens.  
 OS Synthetic.

XX PN WO200069900-A2.

XX PD 23-NOV-2000.

XX PF 17-MAY-2000; 2000WO-US13576.

XX PR 17-MAY-1999; 99US-0134406.

XX PR 10-SEP-1999; 99US-0153406.

XX PR 15-OCT-1999; 99US-0159783.

XX PA (CONJ-) CONJUCHEM INC.

XX PI Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;

XX WPI; 2001-112059/12.

PT Modifying and attaching therapeutic peptides to albumin prevents  
 PT peptidase degradation, useful for increasing length of in vivo activity

XX Disclosure; Page 492; 733pp; English.

XX The present invention describes a modified therapeutic peptide (I)  
 CC comprising a therapeutically active amino acid region (III) and a  
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to  
 CC a less therapeutically active amino acid region (IV), which covalently  
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a  
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.  
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth  
 CC factors and neurotransmitters, to protect them from peptidase activity  
 CC in vivo for the treatment of various disorders. Endogenous therapeutic  
 CC peptides are not suitable as drug candidates as they require frequent  
 CC administration due to rapid degradation by peptidases in the body.

CC Modifying and attaching therapeutic peptides to albumin prevents or  
CC reduces the action of peptidases to increase length of activity (half  
CC life) and specificity as bonding to large molecules decreases  
CC intracellular uptake and interference with physiological processes.  
CC AAB90829 to AAB92441 represent peptides which can be used in the  
CC exemplification of the present invention.  
XX  
SQ Sequence 2 AA;

Query Match 27.3%; Score 6; DB 22; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 2 g 2

RESULT 6  
AAB92149  
ID AAB92149 standard; Peptide; 2 AA.  
XX  
AC AAB92149;  
XX  
DT 22-JUN-2001 (first entry)  
XX  
DE Polypeptide SEQ ID NO:1325.  
XX

Protection; endogenous therapeutic peptide; peptidase; conjugation;  
KW blood component; modification; succinimidyl; maleimido group; amino;  
KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200069900-A2.  
XX  
PD 23-NOV-2000.  
XX  
PF 17-MAY-2000; 2000WO-US13576.  
XX  
PR 17-MAY-1999; 99US-0134406.  
PR 10-SEP-1999; 99US-0153406.  
PR 15-OCT-1999; 99US-0159783.  
XX  
PA (CONJ-) CONJUCHEM INC.  
XX

Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;  
WPI; 2001-112059/12.

Modifying and attaching therapeutic peptides to albumin prevents  
peptidase degradation, useful for increasing length of in vivo activity

Disclosure; Page 630; 733pp; English.

The present invention describes a modified therapeutic peptide (I)  
comprising a therapeutically active amino acid region (III) and a  
reactive group (II) (e.g. succinimidyl and maleimido groups) attached to  
a less therapeutically active amino acid region (IV), which covalently  
bonds with amino/hydroxyl/thiol groups on blood components to form a  
peptidase stabilised therapeutic peptide composed of 3-50 amino acids.  
(I) are useful for modifying therapeutic peptides composed of 3-50 amino acids.  
factors and neurotransmitters, to protect them from peptidase activity  
in vivo for the treatment of various disorders. Endogenous therapeutic  
peptides are not suitable as drug candidates as they require frequent  
administration due to rapid degradation by peptidases in the body.  
Modifying and attaching therapeutic peptides to albumin prevents or  
reduces the action of peptidases to increase length of activity (half  
life) and specificity as bonding to large molecules decreases  
intracellular uptake and interference with physiological processes.

CC AAB90829 to AAB92441 represent peptides which can be used in the  
CC exemplification of the present invention.  
XX  
SQ Sequence 2 AA;

Query Match 27.3%; Score 6; DB 22; Length 2;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
Db 2 g 2

RESULT 7  
AAB92356  
ID AAB92356 standard; Peptide; 2 AA.  
XX  
AC AAB92356;  
XX  
DT 22-JUN-2001 (first entry)  
XX  
DE Miscellaneous peptide SEQ ID NO:1532.

Protection; endogenous therapeutic peptide; peptidase; conjugation;  
KW blood component; modification; succinimidyl; maleimido group; amino;  
KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200069900-A2.  
XX  
PD 23-NOV-2000.  
XX  
PF 17-MAY-2000; 2000WO-US13576.  
XX  
PR 17-MAY-1999; 99US-0134406.  
PR 10-SEP-1999; 99US-0153406.  
PR 15-OCT-1999; 99US-0159783.  
XX  
PA (CONJ-) CONJUCHEM INC.  
XX

Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;  
WPI; 2001-112059/12.

Modifying and attaching therapeutic peptides to albumin prevents  
peptidase degradation, useful for increasing length of in vivo activity

Disclosure; Page 706; 733pp; English.

The present invention describes a modified therapeutic peptide (I)  
comprising a therapeutically active amino acid region (III) and a  
reactive group (II) (e.g. succinimidyl and maleimido groups) attached to  
a less therapeutically active amino acid region (IV), which covalently  
bonds with amino/hydroxyl/thiol groups on blood components to form a  
peptidase stabilised therapeutic peptide composed of 3-50 amino acids.  
(I) are useful for modifying therapeutic peptides composed of 3-50 amino acids.  
factors and neurotransmitters, to protect them from peptidase activity  
in vivo for the treatment of various disorders. Endogenous therapeutic  
peptides are not suitable as drug candidates as they require frequent  
administration due to rapid degradation by peptidases in the body.  
Modifying and attaching therapeutic peptides to albumin prevents or  
reduces the action of peptidases to increase length of activity (half  
life) and specificity as bonding to large molecules decreases  
intracellular uptake and interference with physiological processes.  
AAB90829 to AAB92441 represent peptides which can be used in the  
CC exemplification of the present invention.

XX  
SQ Sequence 2 AA;



```

Query Match      27.3%; Score 6; DB 22; Length 2;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10
Db 2 g 2

RESULT 8
AAP30453
ID AAP30453 standard; peptide; 3 AA.
XX
AC AAP30453;
XX
DT 14-JUN-1992 (first entry)
XX
DE Sequence of collagenase inhibitor.
XX
KW Rheumatoid arthritis therapy; stomach disease; periodontal membrane.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /label= R-G
FT FT /note= "R-H,2-6C alkanoyl,6-8C cycloalkylcarbonyl,
FT FT 2-6C alkoxycarbonyl; pref.
FT FT cyclopentylcarbonyl"
FT Modified-site 4 /label= D,L-V-(3SH)-NH2
FT FT
XX US4371465-A.
XX
PN 01-FEB-1983.
XX
PD
XX 07-OCT-1981; 81US-0309367.
XX
PR 07-OCT-1981; 81US-0309367.
XX
PA (AMHP ) AMERICAN HOME PROD CORP.
XX
PI McGregor WH;
XX
DR WPI; 1983-17252K/07 (17252K).
XX
PS Polypeptide(s) useful as collagenase inhibitors - e.g. in
PT treatment of rheumatoid arthritis, etc.
XX
PS Claim 1; column 4; 3pp; English.
XX
CC The peptides of the invention are collagenase inhibitors useful in
CC the treatment of disease states involving excessive collagen
CC destruction, e.g. rheumatoid arthritis and diseases of the stomach,
CC eye, middle ear, periodontal membranes and skin. AAP30453 gives 86%
CC inhibition of collagenase, c.f. 39% for D-penicillamine in the
CC procedure of Sellers et al., Biochem. J., 167, 353-360, 1977.
XX
SQ Sequence 3 AA;

Query Match      27.3%; Score 6; DB 4; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10
Db 1 g 1

RESULT 9
AAP81081
ID AAP81081 standard; protein; 3 AA.
XX
AC AAP81081;
XX
DT 13-JAN-1991 (first entry)
XX
DE Sequence encoded by the human low density lipoprotein (LDL) receptor
DE promoter.
XX
KW Sterol regulatory element; sterol mediated repression.
XX
OS Homo sapiens.
XX
PN W08807579-A.
XX
PD 06-OCT-1988.
XX
PF 30-MAR-1988; 88WO-US01095.
XX
PR 30-MAR-1987; 87US-0033302.
XX
PA (TEXA ) UNIV OF TEXAS SYST.
XX
PI Brown MS, Goldstein JL, Russell DW, Sudhof TC;
XX
DR WPI; 1988-292863/41.
XX
DR P-PSDB; AAN80195.
XX
DE Sterol regulatory elements and positive promoters -
DE used to control expression of heterologous structural genes and
DE screening plasma cholesterol lowering drugs
XX
PS Example: Fig 1; 99pp; English.
XX
CC A novel purified segment of DNA which comprises a functionally
CC translocatable sterol regulatory element (SRE) confers sterol mediated
CC suppression of structural gene transcription to a selected heterologous
CC structural gene when located upstream from and proximal to a
CC transcription initiation site of such a gene, providing the segment is
CC free of the structural gene ordinarily under the transcriptional control
CC of the SRE is claimed. Also claimed a purified segment of DNA
CC comprising a functional translocatable positive LDL receptor gene
CC promoter control element. The example reflects experiments conducted to
CC display generally the positive capability of the 5'-regions of human LDL
CC receptor. First base in AAN80195 is designated -240 in Fig 1.
XX
SQ Sequence 3 AA;

Query Match      27.3%; Score 6; DB 9; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 G 10
Db 2 g 2

RESULT 10
AAP91660
ID AAP91660 standard; protein; 3 AA.
XX
AC AAP91660;
XX
DT 29-JUN-1990 (first entry)
XX
DE Synthetic peptide corresp. to residues 12-14 of naturally occurring
DE epidermal growth factor (EGF).
XX
KW Epidermal growth factor; angiogenesis; synthetic peptide.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /label=OTHER
FT FT

```

FT Misc-difference /note="H-Gly"  
 FT 3 /label=OTHER  
 FT /note="(aceto amido methyl) NH2-Cys"

PN WO8901489-A.

PD 23-FEB-1989.

PF 10-AUG-1988; 88WO-AU00300.

PR 10-AUG-1987; 87AU-0003629.

PA (CSIR ) COMMONWEALTH SCIENT ORG.

PI McAuslan BR;

DR WPI; 1989-068852/09.

XX Synthetic peptide active in stimulating angiogenesis -  
 PT has sequences corresponding to amino acid sequences occurring in  
 PT epidermal growth factor.

PS Claim 3; page 10; 11pp; English.

XX The inventors claim synthetic peptides which correspond to sequences  
 CC occurring in EGF, but excluding EGF. The peptides are angiogenic  
 CC and have corresp. applications, eg for the healing of wounds and  
 CC burns. Their relative shortness means that they pose fewer synthesis  
 CC problems than the entire EGF molecule. They can be admin. singly or  
 CC association with each other or in association with an angiogenic  
 CC stimulator.

XX Sequence 3 AA;

Query Match 27.3%; Score 6; DB 10; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

DB 1 9 1

RESULT 11

AAR04607  
 ID AAR04607 standard; peptide; 3 AA.

AC AAR04607;

XX 05-SEP-1990 (first entry)

DT Antiviral agent.

DE Antiviral; M2; poliovirus; polio; hepatitis.

KW Synthetic.

XX JP02078631-A.

PN 19-MAR-1990.

PD 14-SEP-1988; 88JP-0228843.

PF 14-SEP-1988; 88JP-0228843.

PR (NIHA ) NIPPON MINING KK.

PA WPI; 1990-129060/17.

XX Antiviral agent contg. tripeptide (unit) -

PT of basic aminoacid, then alanine, glycine or sarcosine, and

PT acidic aminoacid, effective against virus with protein-terminated DNA  
 PT or RNA.

XX Disclosure; ; 4pp; Japanese.

XX Peptide is effective against inhibiting propagation of DNA or RNA  
 CC bonded, protein containing viruses eg. Poliovirus, Hepatitis virus.

XX Sequence 3 AA;

Query Match 27.3%; Score 6; DB 11; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

DB 2 9 2

RESULT 12

AAR00718  
 ID AAR00718 standard; peptide; 3 AA.

AC AAR00718;

XX 29-MAY-1990 (first entry)

DT Core repeat of cell-adhesive protein.

DE Cell adhesion; anti-metastatic agent; immunomodulation; core repeat.

XX EP347931-A.

XX 27-DEC-1989.

XX 23-JUN-1989; 89EP-0111468.

XX 24-JUN-1988; 88JP-0156133.

XX (AZUM/) AZUMA I.

XX Saiki I, Nishi N, Azuma I, Tokura S;

XX WPI; 1990-001305/01.

XX Polypeptide with repeated sequences of cell adhesion protein used  
 PT as anti-metastatic agent for cancer and agonist or antagonist of  
 PT cell-adhesion proteins

PS Claim 2; page 14; 16pp; English.

XX Peptide core is repeated 2-20 times to form a cell-adhesive protein of  
 CC mol. wt. 1,500-5,000. The protein is an (ant)agonist of cell-adhesive  
 CC proteins such as fibronectin. It has high antimetastatic activity  
 CC against cancer and can be used in immunomodulation, wound healing,  
 CC platelet aggregation inhibition and alleviation of neuro-disorders.  
 CC See also AAR00722.

XX Sequence 3 AA;

Query Match 27.3%; Score 6; DB 11; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10

DB 2 9 2

RESULT 13

AAR23219

ID AAR23219 standard; Peptide; 3 AA.  
 XX AC AAR23219;  
 XX DT 18-NOV-1992 (first entry)  
 XX DE Biocompatible polymer-conjugated bioactive peptide.  
 XX DE Metastatic tumour; wound healing; burn healing; angiogenesis;  
 KW KW promoter:atherosclerosis; isocyanate; polyol; prepolymer.  
 XX OS Synthetic.  
 XX US5091176-A.  
 XX PD 25-FEB-1992.  
 XX PF 24-APR-1990; 90US-0510260.  
 XX PR 24-APR-1990; 90US-0510260.  
 XX PR 02-NOV-1988; 88US-0266445.  
 XX PA (GRAC ) GRACE W R & CO-CONN.  
 XX PI Braatz JA, Heifetz AH;  
 XX WPI; 1992-088348/11.  
 XX PT Bioactive peptide improvement by isocyanate-capped poly:ol  
 PT conjugation - giving prods. which are more potent, have decreased  
 PT immunogenicity and longer circulation half life  
 XX Claim 4; column 18; 12pp; English.  
 XX The invention relates to the enhancement of the biological or  
 CC pharmacological activity of an active peptide by preparing an aq.  
 CC solution of a liquid polymer modified version of the peptide by  
 CC covalently binding the peptide to a biocompatible prepolymer under  
 CC aq.conditions. The prepolymer is a triol or higher polyol made up of  
 CC at least 75% oxethylene monomers, has mol.wt. 7000 - 30,000 and has  
 CC all the OH groups capped with aliphatic or cycloaliphatic  
 CC isocyanates, pref. isophorone diisocyanate or methylene bis(cyclo-  
 CC hexyl diisocyanate). The covalent bond is between an NCO group on  
 CC the prepolymer and an amino, SH, OH or COOH group on the peptide.  
 CC The modified peptides have increased potency, decreased immunogenicity  
 CC and longer circulation half-life than the free peptides.  
 CC Suitable peptides include alkaline phosphatase, haemoglobin, RGD  
 CC peptides and GHK peptides. The modified peptides may be variously useful  
 CC for promotion of wound healing, burn healing and angiogenesis, and for  
 CC inhibition of atherosclerosis or tumour cell metastasis.  
 XX Sequence 3 AA;  
 Query Match 27.3%; Score 6; DB 13; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 10 G 10  
 Db 1 g 1  
 RESULT 14  
 AAR36707  
 ID AAR36707 standard; peptide; 3 AA.  
 XX AC AAR36707;  
 XX DT 26-AUG-1993 (first entry)  
 XX DE Adhesion formation prevention RGD-contg. peptide.  
 XX

KW Tissue repair; peritoneum; surgery; post-surgically; inhibition;  
 KW platelet aggregation; cardiovascular; orthopedic; thoracic;  
 KW ophthalmic; CNS; use.  
 XX Synthetic.  
 XX WO9308818-A.  
 XX PD 13-MAY-1993.  
 XX PF 06-NOV-1992; 92WO-US09494.  
 XX PR 07-NOV-1991; 91US-0789231.  
 XX PA (UYSC-) UNIV SOUTHERN CALIFORNIA.  
 XX PI Dizerega GS, Rodgers KE;  
 XX WPI; 1993-167381/20.  
 XX Prevention of adhesion formation, partic. post-surgically - comprises  
 PT administering a RGD-contg. peptide for a time sufficient to permit  
 PT tissue repair  
 XX Example; Page 18; 22pp; English.  
 XX The sequence is that of an RGD-contg. peptide which is used in a  
 CC method for prevention of adhesion formation for a time sufficient  
 CC to permit tissue repair. The method is used for minimising or  
 CC preventing adhesion formation, partic. in the peritoneum following  
 CC surgery, but also for e.g. cardiovascular, orthopedic, thoracic,  
 CC ophthalmic, CNS and other uses. In addn., the peptide inhibits  
 CC platelet aggregation and does not induce inflammation or trauma  
 CC at the site of administration.  
 XX Sequence 3 AA;  
 Query Match 27.3%; Score 6; DB 14; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 10 G 10  
 Db 2 g 2  
 RESULT 15  
 AAR53144  
 ID AAR53144 standard; peptide; 3 AA.  
 XX AC AAR53144;  
 XX DT 02-JUN-1994 (first entry)  
 XX DE RGD peptide derivative #6.  
 XX KW Drug; organ transplantation; rejection; immune disorder;  
 KW systemic lupus.  
 XX Synthetic.  
 XX Key Location/Qualifiers  
 XX Modified-site 1 /note= "C9H19CO-Arg, C13H27CO-Arg, C15H31CO-Arg,  
 FT CH3-[(CH(CH3))-(CH2)3]3-CH(CH3)-CH2-CO-Arg or  
 FT CH3-[(CH(CH3))-(CH2)3]2-CH(CH3)-CH2-CO-Arg"  
 FT Modified-site 4 /note= "Asp-OH or Asp-NH2"  
 XX JP05255105-A.  
 XX PD 05-OCT-1993.

XX 16-MAR-1992; 92JP-0058460.  
 PF 16-MAR-1992; 92JP-0058460.  
 PR (FUJIF ) FUJIFILM PHOTO FILM CO LTD.  
 XX WPI; 1993-348360/44.  
 XX  
 XX Immuno-control drug for organ transplant rejection etc. - contains  
 PT peptide having arginine, glycine, aspartic acid sequence  
 XX  
 PS Disclosure; Page 3; 11pp; Japanese.  
 XX  
 CC The sequences given in AAR44043-47 and AAR53144 represent examples of  
 CC the claimed RGB containing peptide of the invention. These peptides all  
 CC correspond to the generic formulae:  
 CC HO2-(CH2)m-C(O)-([X]-Arg-Gly-Asp-[Y])n-O-CH2CH(OR1)CH2OR2 or  
 CC R3-([X]-Arg-Gly-Asp-[Y])n-Z  
 CC [X], [Y] = amino acid or peptide residues;  
 CC m = 1-5;  
 CC n = 1-5;  
 CC R1, R2 = H or 8-24C acyl or alkyl;  
 CC R3 = 6-24C acyl;  
 CC Z = hydroxyl or amino.  
 CC These peptides form the active part of drugs which are used for the  
 CC control of organ transplantation rejection or immune disorders such  
 CC as systemic lupus.  
 XX  
 SQ Sequence 3 AA;

Query Match 27.3%; Score 6; DB 14; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 G 10  
 Db 2 g 2

Search completed: February 12, 2002, 12:30:34  
 Job time: 367 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:36 ; Search time 232.64 Seconds  
(without alignments)  
11.317 Million cell updates/sec

Title: US-09-485-571-19  
Perfect score: 89  
Sequence: 1 RGRRLAYRLLRFAIRVGR 18  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_17.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_invertebrate.\*
  - 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	45	50.6	648	2	Q9HUH9	Q9huH9 pseudomonas
2	44	49.4	173	2	Q9ETA8	Q9eta8 corynebacte
3	44	49.4	297	12	Q9YNB1	Q9ynb1 sugarcane s
4	44	49.4	588	12	O70681	O70681 sugarcane s
5	44	49.4	1408	1	O27557	O27557 methanobact
6	43	48.3	785	1	Q9P9M2	Q9p9m2 pyrobaculum
7	42	47.2	212	2	Q9R6F9	Q9r6f9 agrobacteri
8	42	47.2	236	2	Q9RYI5	Q9ryi5 deinococcus
9	42	47.2	424	2	Q9F5C0	Q9f5c0 agrobacteri
10	42	47.2	664	2	Q9HVT9	Q9hvt9 pseudomonas
11	42	47.2	679	2	Q9I474	Q9i474 pseudomonas
12	42	47.2	2209	12	Q9YZ25	Q9yz25 lymphocytic
13	41	46.1	447	2	Q9Z571	Q9z571 streptomyce
14	41	46.1	635	2	Q9CDB2	Q9cdb2 mycobacteri
15	41	46.1	637	2	O53580	O53580 mycobacteri
16	40	44.9	207	12	Q9QAR7	Q9qar7 bovine coro
17	40	44.9	207	12	Q9QAR0	Q9qar0 bovine coro
18	40	44.9	207	12	Q9QAQ4	Q9qaq4 bovine coro
19	40	44.9	259	10	Q9LWR8	Q9lwr8 oryza sativ

20	40	44.9	334	2	P73476	P73476 synechocyst
21	40	44.9	368	2	Q9KXR7	Q9kxr7 streptomyce
22	40	44.9	521	2	Q9CLA0	Q9cla0 pasteurilla
23	40	44.9	746	4	Q9BU60	Q9bu60 homo sapien
24	40	44.9	769	2	Q9IOC8	Q9ioc8 pseudomonas
25	40	44.9	1021	4	O15451	O15451 homo sapien
26	40	44.9	1251	4	O15450	O15450 homo sapien
27	39	43.8	198	1	Q9U2U5	Q9uzU5 pyrococcus
28	39	43.8	270	2	Q9CD86	Q9cd86 mycobacteri
29	39	43.8	372	5	O45739	O45739 caenorhabdi
30	39	43.8	447	10	Q9SRM7	Q9srM7 arabidopsis
31	39	43.8	546	2	Q9A910	Q9a910 caulobacter
32	39	43.8	861	2	O06944	O06944 synechocyst
33	39	43.8	1516	3	Q9Y752	Q9y752 candida alb
34	38	42.7	144	8	Q35224	Q35224 oenothera b
35	38	42.7	147	1	Q9UY64	Q9uy64 pyrococcus
36	38	42.7	165	1	O58600	O58600 pyrococcus
37	38	42.7	171	8	Q35645	Q35645 petunia sp.
38	38	42.7	171	8	O99880	O99880 solanum tub
39	38	42.7	175	2	O9F9T9	O9f9t9 edta-degrad
40	38	42.7	224	2	O83422	O83422 treponema p
41	38	42.7	261	2	O9CGJ1	O9cgj1 lactococcus
42	38	42.7	302	2	O07230	O07230 mycobacteri
43	38	42.7	303	2	O9L9P8	O9l9p8 methylobaci
44	38	42.7	357	2	Q9CMC7	Q9cmC7 pasteurilla
45	38	42.7	360	10	Q43705	Q43705 zea mays (m

ALIGNMENTS

RESULT 1

ID Q9HUH9 PRELIMINARY; PRT; 648 AA.

AC Q9HUH9;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE PROBABLE OXIDOREDUCTASE.

GN PA4986.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964(2000).

CC -1- COFACTOR: FAD (BY SIMILARITY).

DR EMBL: AE004911; AAC08371.1; -.

DR InterPro: IPR001327; FAD\_pyr\_redox.

DR InterPro: IPR001155; Oxidored\_FMN.

DR InterPro: IPR001100; pyr\_redox.

DR Pfam: PF00724; oxidored\_FMN; 1.

DR PRINTS: PR00368; FADPNR.

DR PRINTS: PR00411; PNDRTASEI.

KW Complete proteome; FAD; Flavoprotein; Oxidoreductase.

SQ SEQUENCE 648 AA; 70210 MW; ADID78ACF7887D4F CRC64;

Query Match 50.6% Score 45; DB 2; Length 648;  
Best Local Similarity 50.0%; Pred. No. 38;  
Matches 11; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

QY 1 RGGRL----AYRLRLFAIRVGR 18  
| | | | | : : : | | : | | | |  
Db 487 RGGELQVVDQWQLRGEVRVGR 508

RESULT 2  
Q9ETA8  
ID Q9ETA8 PRELIMINARY; PRT; 173 AA.  
AC Q9ETA8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 18.9 KDA PROTEIN.  
GN ORF58.  
OS Corynebacterium equii (Rhodococcus equi).  
OG Plasmid PREAT701 (P33701), and Plasmid virulence.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcu.  
OX NCBI\_TaxID=43767;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC33701; PLASMID-PREAT701 (P33701);  
RA Takai S., Sekizaki T., Kakuda T., Nakamura M., Suzuki K., Ogino N.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC33701, AND 103; PLASMID-PREAT701 (P33701), AND VIRULENCE;  
RX PubMed11083803;  
RA Takai S., Hines S.A., Sekizaki T., Nicholson V.M., Alperin D.A.,  
RA Osaki M., Takamatsu D., Nakamura M., Suzuki K., Ogino N., Kakuda T.,  
RA Dan H., Prescott J.F.;  
RT "DNA sequence and comparison of virulence plasmids from Rhodococcus  
equi ATCC 33701 and 103".  
RL Infect. Immun. 68:6840-6847(2000).  
DR EMBL; AF001204; BAB16667.1; -;  
DR EMBL; AF116907; AAG21761.1; -;  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 173 AA; 18851 MW; F18A637BCA404053 CRC64;

Query Match 49.4%; Score 44; DB 2; Length 173;  
Best Local Similarity 47.1%; Pred. No. 13;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGRLAYRLFAIRVGR 18  
| | | | | : : : | | : | | | |  
Db 142 GGRISYASLYTDTGR 158

RESULT 3  
Q9YNB1  
ID Q9YNB1 PRELIMINARY; PRT; 297 AA.  
AC Q9YNB1;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE NIB (FRAGMENT).  
OS Sugarcane streak mosaic virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
OC Potyvirus.  
OX NCBI\_TaxID=53954;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hena M., Greenivasulu J., Joseph J., Gopinath K., Savithri H.S.;  
RT "Molecular Characterization and intervir relationships of a flexuous  
filamentous virus causing mosaic disease of sugarcane (Saccharum  
officinarium L.) in India".  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y17738; CAA76842.1; -;  
DR InterPro; IPR001592; Poty-coat.  
DR Pfam; PF00767; Poty-coat; 1.  
FT NON\_TER  
SQ SEQUENCE 297 AA; 32843 MW; FE8B424F422A8D15 CRC64;

Query Match 49.4%; Score 44; DB 12; Length 297;  
Best Local Similarity 43.8%; Pred. No. 24;  
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 GRLAYRLFAIRVGR 18  
| | | | | : : : | | : | | | |  
Db 191 GELAYKWQFSVRSKG 206

RESULT 4  
O70681  
ID O70681 PRELIMINARY; PRT; 588 AA.  
AC O70681;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE COAT PROTEIN (FRAGMENT).  
OS Sugarcane streak mosaic virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
OC Potyvirus.  
OX NCBI\_TaxID=53954;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hall J.S., Adams B., Parsons T.J., French R., Lane L., Jensen S.G.;  
RL Mol. Phylogenet. Evol. 0:0-0(1998).  
DR EMBL; U75456; AAC16271.1; -;  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR001592; Poty-coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00767; Poty-coat; 1.  
KW Coat protein.  
FT NON\_TER  
SQ SEQUENCE 588 AA; 66335 MW; 3876EEF59189EA32 CRC64;

Query Match 49.4%; Score 44; DB 12; Length 588;  
Best Local Similarity 43.8%; Pred. No. 50;  
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 GRLAYRLFAIRVGR 18  
| | | | | : : : | | : | | | |  
Db 482 GELAYKWQFSVRSKG 497

RESULT 5  
O27557  
ID O27557 PRELIMINARY; PRT; 1408 AA.  
AC O27557;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CELL SURFACE GLYCOPROTEIN (S-LAYER PROTEIN) RELATED PROTEIN (S-LAYER  
PROTEIN).  
GN MPH1513.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DELTA H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,  
RA Harridge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiweni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
deltaH: functional analysis and comparative genomics.";

RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL; AE000911; AAB85988.1; -  
KW Complete proteome.  
SQ SEQUENCE 1408 AA; 152736 MW; 294EC7742ABB29F2 CRC64;

Query Match 49.4%; Score 44; DB 1; Length 1408;  
Best Local Similarity 41.2%; Pred. No. 1.3e+02;  
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGRLAYRLRFAIRVGR 18  
| : | | | : | : | |  
Db 1318 GAKLGYRTFRFTLKPGR 1334

RESULT 6  
Q9P9M2 PRELIMINARY; PRT; 785 AA.  
AC Q9P9M2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE FAMILY B DNA POLYMERASE.  
GN POLB3.  
OS Pyrobaculum islandicum.  
OC Archaea; Crenarchaeota; Thermoproteales; Thermoproteaceae;  
OC Pyrobaculum.  
OX NCBI\_TaxID=2277;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Geo 3;  
RX MEDLINE=20100754; PubMed=10633098;  
RA Kaehler M., Antranikian G.;  
RT "Cloning and Characterization of a Family B DNA Polymerase from the  
RT Hyperthermophilic Crenarchaeon Pyrobaculum islandicum.";  
RL J. Bacteriol. 182:655-663(2000).  
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N  
CC PYROPHOSPHATE + DNA(N).  
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
DR EMBL; AF195019; AAF27815.1; -  
DR InterPro; IPR002064; DNA\_pol\_B.  
DR Pfam; PF00136; DNA\_pol\_B; 1.  
DR PRINTS; PR00106; DNAPOLB.  
DR SMART; SM00486; POLBc; 1.  
DR PROSITE; PS01116; DNA\_POLYMERASE\_B; 1.  
KW DNA replication; DNA-binding; DNA-directed DNA polymerase.  
SQ SEQUENCE 785 AA; 89748 MW; 10FB8B66A8F3730D CRC64;

Query Match 48.3%; Score 43; DB 1; Length 785;  
Best Local Similarity 50.0%; Pred. No. 1e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLAYRLRFAIRVGR 17  
| : : | | | : | |  
Db 373 GNRVWMLLRAYRLG 388

RESULT 7  
Q9R6F9 PRELIMINARY; PRT; 212 AA.  
AC Q9R6F9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE TIORF100 PROTEIN.  
GN TIORF100.  
OS Agrobacterium radiobacter.  
OG Plasmid pTi-SAKURA.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=358;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF301001;  
RX MEDLINE=20184752; PubMed=10721727;  
RA Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,  
RA Katoh A., Yoshida K.;  
RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";  
RL Gene 247:331-336(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF301001;  
RX MEDLINE=98193120; PubMed=9524202;  
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;  
RT "Novel structural difference between nopaline- and octopline- type trbJ  
RT gene:construction of genetic and physical map and sequencing of  
RT trb/tra and rep gene clusters of a new Ti plasmid pTi-SAKURA.";  
RL Biochim. Biophys. Acta 1396:1-7(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF301001;  
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;  
RT "Genome structure of pTi-SAKURA(I): Strategy for DNA sequencing of a  
RT Japanese cherry-Ti plasmid.";  
RL Nucleic Acids Symp. Ser. 37:159-160(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF301001;  
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;  
RT "Genome structure of pTi-SAKURA (III): Characteristics of T-DNA.";  
RL Nucleic Acids Symp. Ser. 39:185-186(1998).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF301001;  
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;  
RT "Genome structure of pTi-SAKURA (IV): Characteristics of tra region.";  
RL Nucleic Acids Symp. Ser. 39:187-188(1998).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF301001;  
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;  
RT "Genome structure of pTi-SAKURA (V): Complete nucleotide sequence of  
RT plasmid pTi-SAKURA's vir region in Agrobacterium tumefaciens.";  
RL Nucleic Acids Symp. Ser. 39:265-266(1998).  
DR EMBL; AB016260; BAA87725.1; -  
KW Plasmid.  
SQ SEQUENCE 212 AA; 22823 MW; 9533E5438CAC0396 CRC64;

Query Match 47.2%; Score 42; DB 2; Length 212;  
Best Local Similarity 58.3%; Pred. No. 35;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGRLAYRLRFA 13  
| | | | | | | | | |  
Db 116 GGRMEWLLRFS 127

RESULT 8  
Q9RYI5 PRELIMINARY; PRT; 236 AA.  
AC Q9RYI5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL 27.0 KDA PROTEIN.  
GN DRA0327.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.W.;  
 RT "genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1";  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AE001863; AAF12511.1; -;  
 DR TIGR: DRA0327; -;  
 DR InterPro: IPR000051; SAM\_bind.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 236 AA; 27048 MW; 58AA84E95A5DC54C CRC64;

Query Match 47.2%; Score 42; DB 2; Length 236;  
 Best Local Similarity 62.5%; Pred. No. 39;  
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 GGRLAYRLRLFAIRVG 17  
 |||||  
 Db 73 GGLARQLLEWAARDG 88

RESULT 9  
 Q9F5C0 PRELIMINARY; PRT; 424 AA.  
 AC Q9F5C0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE R1ORF132 PROTEIN.  
 GN R1ORF132.  
 OS Agrobacterium rhizogenes.  
 OG plasmid pRi1724.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,  
 RA Yoshida K.;  
 RT "The complete nucleotide sequence of a Ri (root inducing) plasmid  
 RT indicates its chimerical structure between Ti and Sym plasmids.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;  
 RT "Analysis of unique variable region of a plant root inducing plasmid,  
 RT pRi1724, by the construction of its physical map and library.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;  
 RT "Genome structure of Ri plasmid (1): Construction of linking library  
 RT and physical map of pRi1724 in Japanese Agrobacterium.";  
 RL Nucleic Acids Symp. Ser. 39:189-190(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,  
 RA Yoshida K.;  
 RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and  
 RT its flanking regions of pRi1724 in Japanese Agrobacterium  
 RT rhizogenes.";  
 RL Nucleic Acids Symp. Ser. 42:67-68(1999).  
 DR EMBL: AP002086; BAB16251.1; -;  
 DR InterPro: IPR000707; Para.

DR Pfam: PF00991; Para; 1.  
 KW Plasmid.  
 SQ SEQUENCE 424 AA; 47098 MW; 24AF2FB68BFC8A87 CRC64;

Query Match 47.2%; Score 42; DB 2; Length 424;  
 Best Local Similarity 69.2%; Pred. No. 75;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 GGRLAYRLRLFEAI 14  
 |||||  
 Db 329 GGRLDYDFLRLFI 341

RESULT 10  
 Q9HYT9 PRELIMINARY; PRT; 664 AA.  
 AC Q9HYT9;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HYPOTHETICAL PROTEIN PA3305.  
 GN PA3305  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RA MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuen Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 RL EMBL: AE004753; AAG06693.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 664 AA; 73136 MW; 19EBBF3098D1F4E CRC64;

Query Match 47.2%; Score 42; DB 2; Length 664;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 GGRLAYRLRLRFA 13  
 |||||  
 Db 536 GGRNADRLRLRA 547

RESULT 11  
 Q91474 PRELIMINARY; PRT; 679 AA.  
 AC Q91474;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE HYPOTHETICAL PROTEIN PA1270.  
 GN PA1270.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RA MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,



```
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AF004556; AAG04659.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 679 AA; 73079 MW; B37B15EEED4E5DC7 CRC64;

Query Match 47.2%; Score 42; DB 2; Length 679;
Best Local Similarity 58.9%; Pred. No. 1.3e+02;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGRLAYRLRLFAIRVGR 18
   ||||| ||||| ||
Db 539 GGRTADRLRLAQRYDR 555

RESULT 12
Q9YZ25 PRELIMINARY; PRT; 2209 AA.
AC Q9YZ25;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE L PROTEIN.
OS Lymphocytic choriomeningitis virus (strain WE).
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.
OX NCBI_TaxID=11627;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WE;
RX MEDLINE=99073699; PubMed=9857988;
RA Djavani M., Lukashevich I.S., Salvato M.S.;
RT "Sequence comparison of the large genomic RNA segments of two strains
RT of lymphocytic choriomeningitis virus differing in pathogenic
RT potential for guinea pigs.";
RL Virus Genes 17:151-155(1998).
DR EMBL; AF004519; AAD03396.1; -.
SQ SEQUENCE 2209 AA; 254432 MW; 40104EFA50EDCCCC CRC64;

Query Match 47.2%; Score 42; DB 12; Length 2209;
Best Local Similarity 66.7%; Pred. No. 4.7e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 RLAYRLRLFAIR 15
   :: ||||| ||
Db 662 KVVYRLRLFLIR 673

RESULT 13
Q9Z571 PRELIMINARY; PRT; 447 AA.
AC Q9Z571;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SC8D9.20C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL035569; CAB37584.1; -.
SQ SEQUENCE 447 AA; 46718 MW; 01D72AC1DC201A5B CRC64;

Query Match 46.1%; Score 41; DB 2; Length 447;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGSRLAYRLRLFAIR 15
   ||||| |||||
Db 16 RGSRLAYRSGRFAAR 30

RESULT 14
Q9CDB2 PRELIMINARY; PRT; 635 AA.
AC Q9CDB2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE ACYL-COA SYNTHETASE.
GN FADD32 OR MLO100.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Bigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL583917; CAC29608.1; -.
KW Complete proteome.
SQ SEQUENCE 635 AA; 69534 MW; C5B15CAD365A46BE CRC64;

Query Match 46.1%; Score 41; DB 2; Length 635;
Best Local Similarity 61.5%; Pred. No. 1.7e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGSRLAYRLRLFA 13
   || :||| | |
Db 38 RGDKLAYRFLDFS 50

RESULT 15
O53580 PRELIMINARY; PRT; 637 AA.
ID O53580
AC O53580;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
```

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE PUTATIVE POLYPEPTIDE SYNTHASE.  
 GN FADD32 OR RV3801C OR MTV026.06C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Suiston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 DR EMBL; AL022076; CAA17865.1; -.  
 DR Tuberculist; RV3801C; -.  
 DR InterPro; IPR000873; AMP-bind.  
 DR Pfam; PF00501; AMP-binding; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 637 AA; 69231 MW; 0A3A86CED9AE0EDC CRC64;

Query Match 46.1%; Score 41; DB 2; Length 637;  
 Best Local Similarity 61.5%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 RGRRLAYRLRLFA 13  
 || :||||| :  
 Db 40 RGDKLAYRFLDFS 52

Search completed: February 12, 2002, 12:38:37  
 Job time: 750 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:49 ; Search time 67.2 Seconds  
(without alignments)  
9.821 Million cell updates/sec

Title: US-09-485-571-19

Perfect score: 89

Sequence: 1 RGGRLAYRLLRFAIRVGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36564827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	52	58.4	149	1	PG1_PIG
2	51	57.3	149	1	PG5_PIG
3	48	53.9	147	1	PG2_PIG
4	45	50.6	149	1	PG3_PIG
5	42	47.2	219	1	SFSA_ARCFU
6	40	44.9	207	1	YIOR_CVBF
7	40	44.9	207	1	YIOR_CVBM
8	40	44.9	207	1	YIOR_CVTK
9	39	43.8	298	1	TRUE_MYCTU
10	39	43.8	356	1	Y42B_RHISN
11	39	43.8	362	1	ALEU_HORVU
12	39	43.8	630	1	GIDA_PSEPU
13	39	43.8	1220	1	DP3A_THEAQ
14	38	42.7	155	1	RM1C_PETHY
15	38	42.7	177	1	RM1C_BRANA
16	38	42.7	185	1	RM16_MAIZE
17	38	42.7	185	1	RM16_ORYSA
18	38	42.7	283	1	NAT_MYCTU
19	38	42.7	315	1	MIAA_VIBCH
20	38	42.7	360	1	CYS2_MAIZE
21	38	42.7	623	1	Y711_HUMAN
22	38	42.7	705	1	MLLC_STRCO
23	38	42.7	784	1	DPO2_AERPE
24	37.5	42.1	488	1	MM11_HUMAN
25	37.5	42.1	492	1	MM11_MOUSE
26	37	41.6	188	1	C561_ECOLI
27	37	41.6	266	1	YK11_CAEEL
28	37	41.6	427	1	HEM1_SYNY3
29	37	41.6	591	1	DS01_PSEAE
30	37	41.6	1661	1	YBTL_YEAST
31	37	41.6	1876	1	PPSA_MYCTU
32	36.5	41.0	136	1	R141_YEAST
33	36.5	41.0	137	1	R142_YEAST

#### ALIGNMENTS

RESULT 1  
PG1\_PIG  
ID PG1\_PIG STANDARD; PRT; 149 AA.  
AC P32194;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PROTEGRIN 1 PRECURSOR (PG-1) (NEUTROPHIL PEPTIDE 1).  
GN NPG1.

OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID:9823;

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=94283613; PubMed=8013647;

RA Zhao C., Liu L., Lehrer R.I.;  
RT "Identification of a new member of the protegrin family by cDNA  
cloning.";

RL FEBS Lett. 346:285-288(1994).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=3ED DUROC;  
RX MEDLINE=95354835; PubMed=7628604;

RA Zhao C., Ganz T., Lehrer R.I.;  
RT "The structure of porcine protegrin genes.";

RL FEBS Lett. 368:197-202(1995).  
RN [3]

RP SEQUENCE OF 131-148.  
RC TISSUE=Leukocyte;

RX MEDLINE=93327946; PubMed=8335113;  
RA Kokryakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,  
Aleshina G.V., Shamova O.V., Korneva H.A., Lehrer R.I.;

RT "Protegrins: leukocyte antimicrobial peptides that combine features  
of corticostatic defensins and tachyplesins.";  
RL FEBS Lett. 327:231-236(1993).  
RN [4]

RP SEQUENCE OF 131-148.  
RC TISSUE=Neutrophils;

RX MEDLINE=93387466; PubMed=8375505;  
RA Mirgorodskaya O.A., Shevchenko A.A., Abdalla K.O.M.A.,  
Chernushevich I.V., Egorov T.A., Musoliamov A.X., Kokryakov V.N.,  
Shamova O.V.;

RT "Primary structure of three cationic peptides from porcine  
neutrophils. Sequence determination by the combined usage of  
electrospray ionization mass spectrometry and Edman degradation.";

RL FEBS Lett. 330:339-342(1993).  
RN [5]

RP STRUCTURE BY NMR OF PROTEGRIN 1.  
RX MEDLINE=96235220; PubMed=8647100;

RA Amelase A., Mangoni M., Roumestand C., Chiche L., Despaux E.,  
Grassy G., Calas B., Chavanieu A.;

RT "Synthesis and solution structure of the antimicrobial peptide  
protegrin-1.";

RL Eur. J. Biochem. 237:575-583(1996).

P27069 kluyveromyc  
P37685 escherichia  
P41014 bacillus ca  
P09221 bacillus ps  
Q00110 ictalurid h  
Q9yfn1 aeropyrum p  
P25778 oryza sativ  
P45079 haemophilus  
P48301 mus musculus  
O27509 methanobact  
P29521 daucus caro  
Q01447 fusarium so

34 36.5 41.0 137 1 RS14\_KLULA  
35 36.5 41.0 512 1 ALDB\_ECOLI  
36 36 40.4 162 1 ATPF\_BACCA  
37 36 40.4 163 1 ATPF\_BACP3  
38 36 40.4 227 1 VG70\_HSV11  
39 36 40.4 238 1 RL2\_AERPE  
40 36 40.4 362 1 ORYC\_ORYSA  
41 36 40.4 440 1 YDJN\_HABIN  
42 36 40.4 445 1 TEFA\_MOUSE  
43 36 40.4 447 1 COBB\_METH  
44 36 40.4 449 1 EF11\_DAUCA  
45 36 40.4 485 1 ER24\_FUSSO

RN  
 RP STRUCTURE BY NMR OF PROTEGRIN 1.  
 RX MEDLINE=97113279; PubMed=8807886;  
 RA Fahrner R.L., Dieckmann T., Harwig S.S., Lehrer R.I., Eisenberg D.,  
 RA Feigon J.;  
 RT "Solution structure of protegrin-1, a broad-spectrum antimicrobial  
 peptide from porcine leukocytes.";  
 RL Chem. Biol. 3:543-550(1996).  
 CC  
 CC -!- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA  
 CC MONOCYTOGENES AND C.ALBCANS, IN VITRO.  
 CC  
 CC -!- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; X79868; CAA56251.1; -  
 DR EMBL; X84094; CAA58890.1; -  
 DR EMBL; X84094; CAA58890.1; -  
 DR PIR; S34585;  
 DR PIR; S36820; S36820.  
 DR PDB; 1PG1; 27-MAY-98.  
 DR InterPro: IPR001894; Cathelicidin.  
 DR Pfam; PF00666; Cathelicidins; 1.  
 DR ProDom; PD001838; Cathelicidins; 1.  
 DR PROSITE; PS00946; CATHELICIDINS\_1; 1.  
 DR PROSITE; PS00947; CATHELICIDINS\_2; 1.  
 KW Antibiatic; Signal; Amidation; Multigene family; 3D-structure.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 130  
 FT CHAIN 131 148  
 FT MOD\_RES 30 30  
 FT PYRROLIDONE CARBOXYLIC ACID (BY  
 FT SIMILARITY).  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT DISULFID 85 96  
 FT DISULFID 107 124  
 FT DISULFID 136 145  
 FT DISULFID 138 143  
 FT MOD\_RES 148 148  
 FT SEQUENCE 149 AA; 16677 MW; 6EFBA98429CD6E4 CRC64;  
 SQ  
 Query Match 58.4%; Score 52; DB 1; Length 149;  
 Best Local Similarity 61.1%; Pred. No. 0.031;  
 Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 RGGRLAYRLLRFAIRVGR 18  
 Db 131 RGGRLCYCRPRFCVCVGR 148  
 ||||| | | | |  
 RESULT 2  
 PG5\_PIG STANDARD; PRT; 149 AA.  
 ID PG5\_PIG  
 AC P49334;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROTEGRIN 5 PRECURSOR (PG-5).  
 GN NPG5.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RED DUROC;  
 RX MEDLINE=95354835; PubMed=7628604;  
 RA Zhao C., Ganz T., Lehrer R.I.;  
 RT "The structure of porcine protegrin genes.";  
 RL FEBS Lett. 368:197-202(1995).

CC  
 CC -!- FUNCTION: MICROBICIDAL ACTIVITY (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; X84096; CAA58892.1; -  
 DR HSSP; P32194; 1PG1.  
 DR InterPro: IPR001894; Cathelicidin.  
 DR Pfam; PF00666; Cathelicidins; 1.  
 DR ProDom; PD001838; Cathelicidins; 1.  
 DR PROSITE; PS00946; CATHELICIDINS\_1; 1.  
 DR PROSITE; PS00947; CATHELICIDINS\_2; 1.  
 KW Antibiatic; Amidation; Multigene family; Signal.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 130  
 FT CHAIN 131 148  
 FT MOD\_RES 30 30  
 FT PYRROLIDONE CARBOXYLIC ACID (BY  
 FT SIMILARITY).  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT DISULFID 85 96  
 FT DISULFID 107 124  
 FT DISULFID 136 145  
 FT DISULFID 138 143  
 FT MOD\_RES 148 148  
 FT SEQUENCE 149 AA; 16604 MW; 6CC7262429CD6B64 CRC64;  
 SQ  
 Query Match 57.3%; Score 51; DB 1; Length 149;  
 Best Local Similarity 61.1%; Pred. No. 0.046;  
 Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 RGGRLAYRLLRFAIRVGR 18  
 Db 131 RGGRLCYCRPRFCVCVGR 148  
 ||||| | | | |  
 RESULT 3  
 PG2\_PIG STANDARD; PRT; 147 AA.  
 ID PG2\_PIG  
 AC P32195;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROTEGRIN 2 PRECURSOR (PG-2).  
 GN NPG2.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=94071898; PubMed=8250892;  
 RA Storici P., Zanetti M.;  
 RT "A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide  
 with a cathelin-like pro-sequence.";  
 RL Biochem. Biophys. Res. Commun. 196:1363-1368(1993).  
 RN [2]  
 RP SEQUENCE OF 131-146.  
 RC TISSUE=Leukocyte;  
 RX MEDLINE=93327946; PubMed=8335113;  
 RA Kokryakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,  
 RA Aleshina G.M., Shamova O.V., Korneva H.A., Lehrer R.I.;  
 RT "Protegrins: leukocyte antimicrobial peptides that combine features  
 of corticostatic defensins and tachyplesins.";  
 RL FEBS Lett. 327:231-236(1993).  
 CC -!- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA

```

CC MONOCYTOGENES AND C. ALBICANS, IN VITRO.
CC -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L24745; AAA31061.1; -.
DR HSP; P32194; IPI.
DR InterPro: IPR001894; Cathelicidin.
DR Pfam; PF00666; Cathelicidins; 1.
DR ProDom; PD001838; Cathelicidins; 1.
DR ProSITE; PS00946; CATHELICIDINS_1; 1.
DR ProSITE; PS00947; CATHELICIDINS_2; 1.
KW Antibiotic; Signal; Amidation; Multigene family.
FT SIGNAL 1 29
FT PROPEP 30 130
FT CHAIN 131 146
FT MOD_RES 30 30
FT PYRROLIDONE CARBOXYLIC ACID (BY
FT SIMILARITY).
FT DISULFID 85 96
FT DISULFID 107 124
FT DISULFID 136 145
FT DISULFID 138 143
FT MOD_RES 146 146
FT SEQUENCE 147 AA; 16478 MW; 698429DFEC40466 CRC64;

Query Match 53.9%; Score 48; DB 1; Length 147;
Best Local Similarity 64.7%; Pred. No. 0.16;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYLLRFAIRVG 17
Db 131 RGGRLCYCRRRRCICVG 147

RESULT 4
PG3_PIG STANDARD; PRT; 149 AA.
AC P32196;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEGRIN 3 PRECURSOR (PG-3).
GN NPG3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=94283613; PubMed=8013647;
RA Zhao C., Liu L., Lehrer R.I.;
RT "Identification of a new member of the protegrin family by cDNA
RT cloning."
RL FEBS Lett. 346:285-288(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RED DUROC;
RX MEDLINE=95354835; PubMed=7628604;
RA Zhao C., Ganz T., Lehrer R.I.;
RT "The structure of porcine protegrin genes."
RL FEBS Lett. 368:197-202(1995).
RN [3]
RP SEQUENCE OF 131-148.
RC TISSUE=Leukocyte;
RX MEDLINE=93327946; PubMed=8335113;

```

```

RA Kokryakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,
RA Aleshina G.M., Shamova O.V., Korneva H.A., Lehrer R.I.;
RT "Protegrins: leukocyte antimicrobial peptides that combine features
RT of corticostatic defensins and tachyplesins."
RL FEBS Lett. 327:231-236(1993).
CC -1- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA
CC MONOCYTOGENES AND C. ALBICANS, IN VITRO.
CC -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; XH3267; CAA58240.1; -.
DR EMBL; XH4095; CAA58891.1; -.
DR PIR; S34587; S34587.
DR HSP; P32194; IPI.
DR InterPro: IPR001894; Cathelicidin.
DR Pfam; PF00666; Cathelicidins; 1.
DR ProDom; PD001838; Cathelicidins; 1.
DR ProSITE; PS00946; CATHELICIDINS_1; 1.
DR ProSITE; PS00947; CATHELICIDINS_2; 1.
KW Antibiotic; Signal; Amidation; Multigene family.
FT SIGNAL 1 29
FT PROPEP 30 130
FT CHAIN 131 148
FT MOD_RES 30 30
FT PYRROLIDONE CARBOXYLIC ACID (BY
FT SIMILARITY).
FT DISULFID 85 96
FT DISULFID 107 124
FT DISULFID 136 145
FT DISULFID 138 143
FT MOD_RES 148 148
FT SEQUENCE 149 AA; 16578 MW; 6F4BA98429CD6ED4 CRC64;

Query Match 50.6%; Score 45; DB 1; Length 149;
Best Local Similarity 55.6%; Pred. No. 0.54;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGGRLAYLLRFAIRVGR 18
Db 131 RGGRLCYCRRRRCVCVGR 148

RESULT 5
SFSA_ARCFU STANDARD; PRT; 219 AA.
ID SFSA_ARCFU
AC O28756;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SUGAR FERMENTATION STIMULATION PROTEIN HOMOLOG.
GN SFSA OR AF1516
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk E.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,

```

RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 reducing archaeon *Archaeoglobus fulgidus*.";  
 RL Nature 390:364-370(1997).  
 CC -1- SIMILARITY: BELONGS TO THE SFSA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000997; AAB89729.1; -;  
 DR TIGR; AF1516; -;  
 KW Complete proteome.  
 SQ SEQUENCE 219 AA; 25061 MW; EAD775628C39F915 CRC64;

Query Match 47.2%; Score 42; DB 1; Length 219;  
 Best Local Similarity 64.3%; Pred. No. 2.7;  
 Matches 9; Conservative 3; Mismatches 0; Indels 2; Gaps 1;  
 QY 1 RGGRLAYRLLRFAI 14  
 |||:|:|:|:|:|  
 Db 58 RGGKLSYRL--FAV 69

RESULT 6  
 YIOR\_CVBF STANDARD; PRT; 207 AA.  
 AC P22654;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DE 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE HYPOTHETICAL PROTEIN IN NUCLEOCAPSID ORF (IORF).  
 OS Bovine coronavirus (strain F15).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11129;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89087718; PubMed=3207501;  
 RA Cruciere C., Laporte J.;  
 RT "Sequence and analysis of bovine enteric coronavirus (F15) genome.  
 RT I. Sequence of the gene coding for the nucleocapsid protein; analysis  
 RT of the predicted protein.";  
 RL Ann. Inst. Pasteur Virol. 139:123-138(1988).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M36656; AAA42759.1; -;  
 DR PIR; S06869; S06869.  
 KW Hypothetical protein.  
 SQ SEQUENCE 207 AA; 23001 MW; A4E5DE61171BAB50 CRC64;

Query Match 44.9%; Score 40; DB 1; Length 207;  
 Best Local Similarity 53.8%; Pred. No. 5.7;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 RLAYRLLRFAIRV 16  
 ||:|:|:|:|:|

Db 153 RLGRFLARYSLRV 165

RESULT 7  
 YIOR\_CVBM STANDARD; PRT; 207 AA.  
 AC P10525;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE HYPOTHETICAL PROTEIN IN NUCLEOCAPSID ORF (IORF).  
 OS Bovine coronavirus (strain Mebus).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11132;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87151119; PubMed=3029965;  
 RA Lapps W.E., Hogue B.G., Brian D.A.;  
 RT "Sequence analysis of the bovine coronavirus nucleocapsid and matrix  
 RT protein genes.";  
 RL Virology 157:47-57(1987).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M16620; AAA66398.1; -;  
 DR PIR; C26347; Q0IHBC.  
 KW Hypothetical protein.  
 SQ SEQUENCE 207 AA; 23054 MW; BE76DC4D63DD32A CRC64;

Query Match 44.9%; Score 40; DB 1; Length 207;  
 Best Local Similarity 53.8%; Pred. No. 5.7;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 RLAYRLLRFAIRV 16  
 ||:|:|:|:|:|

Db 153 RLGRFLARYSLRV 165

RESULT 8  
 YIOR\_CVTKE STANDARD; PRT; 207 AA.  
 AC P26626;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE HYPOTHETICAL PROTEIN IN NUCLEOCAPSID ORF (IORF).  
 OS Turkey enteric coronavirus (TCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11152;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=MINNESOTA;  
 RX MEDLINE=91311418; PubMed=1856695;  
 RA Verbeek A., Tijssen P.;  
 RT "Sequence analysis of the turkey enteric coronavirus nucleocapsid and  
 RT membrane protein genes: a close genomic relationship with bovine  
 RT coronavirus.";  
 RL J. Gen. Virol. 72:1659-1666(1991).  
 DR PIR; J01174; J01174.  
 KW Hypothetical protein.  
 SQ SEQUENCE 207 AA; 23005 MW; F69E2D8F2F006F77 CRC64;

Query Match 44.9%; Score 40; DB 1; Length 207;

Best Local Similarity 53.8%; Pred. No. 5.7;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 RLAYRLRFAIRV 16  
|| :||| :||| :|||  
Db 153 RLGLRLARYSLRV 165

## RESULT 9

TRUB\_MYCTU STANDARD; PRT; 298 AA.  
AC Q3335;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TRNA PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70) (TRNA PSEUDOURIDINE 5S  
DE SYNTHASE) (PSI55 SYNTHASE) (PSEUDOURIDYLATE SYNTHASE) (URACIL  
DE HYDROLYASE).  
GN TRUB OR RV2793C OR MT2862.1 OR MTV002.58C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Church C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.;  
RL Nature 393:537-544(1998).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE  
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE  
CC 5'-PHOSPHATE + H(2)O  
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: AL008967; CAAL1588.1; -;  
CC TIGR: MT2862.1;  
CC Tuberculist: RV2793c; -;  
DR InterPro: IPR002501; Trub\_N.  
DR Pfam: PF01509; Trub\_N; 1.  
KW Lyase; tRNA processing; Complete proteome.  
SQ SEQUENCE 298 AA; 31819 MW; 9F77797DC13B34C6 CRC64;

Query Match 43.8%; Score 39; DB 1; Length 298;  
Best Local Similarity 80.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGHAYRLRLR 11  
||| |||||  
Db 127 GGHAYRLRLR 136

## RESULT 10

Y4ZB\_RHISN STANDARD; PRT; 356 AA.  
AC P55729;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PUTATIVE TRANSPOSASE Y4ZB.  
GN Y4ZB.  
OS Rhizobium sp. (strain NGR234).  
OG Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97305956; PubMed=9163424;  
RA Freiberg C.A., Fellay R., Broughton W.J., Rosenthal A.,  
RA Perret X.;  
RT "Molecular basis of symbiosis between Rhizobium and legumes";  
RL Nature 387:394-401(1997).  
CC -!- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 11.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: A3000108; AAB91960.1; -;  
DR InterPro: IPR002559; Transposase\_11.  
DR Pfam: PF01609; Transposase\_11; 1.  
KW Hypothetical protein; Transposable element; Transposition;  
KW DNA-binding; DNA recombination; Plasmid.  
SQ SEQUENCE 356 AA; 39775 MW; 9F350ABB7E691635 CRC64;

Query Match 43.8%; Score 39; DB 1; Length 356;  
Best Local Similarity 57.1%; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 LAYRLRFAIRVGR 18  
:||| ||| :|||  
Db 293 IAYALLRIARLNR 306

## RESULT 11

ALEU\_HORVU STANDARD; PRT; 362 AA.  
AC P05167;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE THIOLE PROTEASE ALEURAIN PRECURSOR (EC 3.4.22.16).  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86016732; PubMed=3901004;

```

RA Rogers J.C., Dean D., Heck G.R.;
RT "Aleurain: a barley thiol protease closely related to mammalian
RT cathepsin H.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6512-6516(1985).
RN [2]
RP REVISIONS.
RA Rogers J.C.;
RL Submitted (MAR-1987) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS; ACTS ALSO AS AN
CC AMINOPEPTIDASE ON PEPTIDE SUBSTRATES WITH FREE N-TERMINI.
CC -!- SUBCELLULAR LOCATION: VACUOLE-LIKE SUBCELLULAR COMPARTMENT.
CC -!- INDUCTION: ALEURAIN IS SYNTHESIZED BY THE ALEURONE CELLS
CC STIMULATED BY GIBBERELIC OR ABSICISIC ACID.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC PAPAINE FAMILY OF THIOL PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05167; CAA28804.1; -
DR PIR; A25492; KHBH.
DR HSSP; P07711; ICUL.
DR MEROPS; C01.041; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR00169; Thiolprot_act_site.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAINE.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR HydroLase; Thiol protease; Glycoprotein; Zymogen; Signal; Germination.
KW SIGNAL.
FT SIGNAL 1 ?
FT PROPEP 1 ?
FT CHAIN 144 362
FT ACT_SITE 168 168
FT ACT_SITE 308 308
FT ACT_SITE 328 328
FT DISULFID 165 208
FT DISULFID 199 241
FT DISULFID 299 349
FT CARBOHYD 188 188
FT CARBOHYD 257 257
SQ SEQUENCE 362 AA; 39122 MW; A70CCD4A843A1686 CRC64;

Query Match 43.88; Score 39; DB 1; Length 362;
Best Local Similarity 53.38; Pred. No. 15;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 RLAYRLRFAIRVGR 18
| | | | |
DB 56 RHALRFARFAVRGK 70

RESULT 12
GIDA_PSEPU
ID GIDA_PSEPU STANDARD; PRT; 630 AA.
AC P25756;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUCOSE INHIBITED DIVISION PROTEIN A.
GIDA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=303;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=TN2100;
RX MEDLINE=92204018; PubMed=1552862;
RA Ogasawara N., Yoshikawa H.;
RT "Genes and their organization in the replication origin region of the
RT bacterial chromosome.";
RL Mol. Microbiol. 6:629-634(1992).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62540; CAA44419.1; -
DR PIR; JQ1233; BWPSAP.
DR InterPro; IPR002218; GIDA.
DR Pfam; PF01134; GIDA; 1.
DR ProDom; PD003738; GIDA; 1.
DR PROSITE; PS01280; GIDA_1; 1.
DR PROSITE; PS01281; GIDA_2; 1.
DR PROSITE; PS01281; GIDA_2; 1.
SQ SEQUENCE 630 AA; 69495 MW; 03B86C228F413E7C CRC64;

Query Match 43.88; Score 39; DB 1; Length 630;
Best Local Similarity 40.08; Pred. No. 26;
Matches 10; Conservative 3; Mismatches 4; Indels 8; Gaps 1;

QY 2 GGR-----LAYRLRFAIRVGR 18
| | | | |
DB 172 GGRAGDPPSIALAHMRLEPLRVGR 196

RESULT 13
DP3A_THEAQ
ID DP3A_THEAQ STANDARD; PRT; 1220 AA.
AC G9XDH5;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE III ALPHA SUBUNIT (EC 2.7.7.7).
GN DNAE.
OS Thermus aquaticus.
OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX NCBI_TaxID=271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99246387; PubMed=10229580;
RA Huang Y.P., Ito J.;
RT "DNA polymerase C of the thermophilic bacterium Thermus aquaticus:
RT classification and phylogenetic analysis of the family C DNA
RT polymerases.";
RL J. Mol. Evol. 48:756-769(1999).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLIIV' COMPLEX. POLIIV' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. DNAE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:36 ; Search time 126.85 seconds  
(without alignments)  
10.809 Million cell updates/sec

Title: US-09-485-571-19

Perfect score: 89  
Sequence: 1 RGGRLAYLLRFAIRVGR 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	58.4	149	2 S57607	protegrin 1 precu
2	51	57.3	149	2 S57609	protegrin 5 precu
3	48	53.9	147	2 JN0900	protegrin 2 precu
4	45	50.6	149	2 A53895	protegrin 3 precu
5	45	50.6	648	2 C83023	probable oxidoredu
6	44	49.4	1408	2 H69068	cell surface glyco
7	42	47.2	219	2 C69439	sugar fermentation
8	42	47.2	236	2 D75587	hypothetical prote
9	42	47.2	664	2 D83231	hypothetical prote
10	42	47.2	679	2 A83488	hypothetical prote
11	41	46.1	447	2 T35824	probable oxidoredu
12	41	46.1	637	2 E70887	probable fadD32 pr
13	40	44.9	207	1 QJ1HBC	hypothetical prote
14	40	44.9	207	1 JQ1174	hypothetical prote
15	40	44.9	207	1 S06869	hypothetical prote
16	40	44.9	334	2 S77413	hypothetical prote
17	40	44.9	769	2 B83307	probable molybdopt
18	39	43.8	198	2 D75082	phosphate abc tran
19	39	43.8	298	2 H70884	probable tRNA modi
20	39	43.8	361	1 KBBH	aleurain (PC 3.4.2
21	39	43.8	372	2 T24392	hypothetical prote
22	39	43.8	630	1 BWF5AP	gida protein - pse
23	39	43.8	861	2 S77086	hypothetical prote
24	38	42.7	144	2 S43766	ribosomal protein
25	38	42.7	147	2 F75013	hypothetical prote
26	38	42.7	155	2 S32194	ribosomal protein
27	38	42.7	165	2 B71138	hypothetical prote
28	38	42.7	177	2 S36914	ribosomal protein
29	38	42.7	185	1 R52M6M	ribosomal protein

probable ribosomal  
probable phosphogl  
cobyrilic acid synth  
hypothetical prote  
tRNA delta(2)-isop  
probable nhoA prot  
cysteine proteinase  
probable nuclear p  
nucleoporin homolo  
hypothetical prote  
probable integral  
probable DNA-direc  
poly(A) polymerase  
hypothetical prote  
probable large gly  
inositol 1,4,5-tri

ALIGNMENTS

RESULT 1  
S57607  
protegrin 1 precursor - pig  
N:Alternate names: neutrophil peptide 1  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S66284; S45712; S36820; S34585; S57607  
R:Zhao, C.; Ganz, T.; Lehrer, R.I.  
FEBS Lett. 368, 197-202, 1995  
A:Title: The structure of porcine protegrin genes.  
A:Reference number: S66283; MUID:95354835  
A:Accession: S66284  
A:Molecule type: DNA  
A:Residues: 1-149 <ZHA>  
A:Cross-references: EMBL:X84094; NID:9887642; PIDN:CAA58890.1; PID:9887643  
R:Zhao, C.; Liu, L.; Lehrer, R.I.  
FEBS Lett. 346, 285-288, 1994  
A:Title: Identification of a new member of the protegrin family by cDNA cloning.  
A:Reference number: S45712; MUID:94283613  
A:Accession: S45712  
A:Molecule type: mRNA  
A:Residues: 1-149 <ZH2>  
A:Cross-references: GB:X79868; NID:9603035; PIDN:CAA56251.1; PID:9603036  
R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Eg  
FEBS Lett. 330, 339-342, 1993  
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Seq  
A:Reference number: S36820; MUID:93387466  
A:Accession: S36820  
A:Molecule type: protein  
A:Residues: 131-148 <MIR>  
R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M  
FEBS Lett. 327, 231-236, 1993  
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of cort  
A:Reference number: S34585; MUID:93327946  
A:Accession: S34585  
A:Molecule type: protein  
A:Residues: 131-148 <KOK>  
C:Genetics:  
A:Gene: NPG1  
A:Introns: 66/3; 102/3; 126/3  
C:Superfamily: cathelin; cystatin homology  
C:Keywords: amidated carboxyl end; antibacterial; neutrophil  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:22-129/Domain: cystatin homology <CYS>  
F:30-130/Domain: propeptide #status predicted <PRO>  
F:131-148/Product: protegrin 1 #status experimental <MAT>  
F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followi

Query Match 58.4%; Score 52; DB 2; Length 149;  
Best Local Similarity 61.1%; Pred. No. 0.14;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18  
 ||||| | || : |||  
 Db 131 RGGRLCYCRRRRCVGVGR 148

## RESULT 2

S57609  
 protegrin 5 precursor - pig  
 N:Alternate names: cathelin-associated antimicrobial peptide  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 16-Jul-1999  
 C:Accession: S66283; S57609  
 R:Zhao, C.; Ganz, T.; Lehrer, R.I.  
 FEBS Lett. 368, 197-202, 1995  
 A:Title: The structure of porcine protegrin genes.  
 A:Reference number: S66283; MUID:95354835  
 A:Molecule type: DNA  
 A:Accession: S66283  
 A:Residues: 1-149 <ZHA>  
 A:Cross-references: EMBL:X84096; NID:g887646; PIDN:CAA58892.1; PID:g887647  
 A:Experimental source: leukocytes  
 C:Genetics:  
 A:Gene: NPG5  
 A:Introns: 66/3; 102/3; 126/3  
 C:Superfamily: cathelin; cystatin homology  
 C:Keywords: amidated carboxyl end; antibacterial  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 F:22-129/Domain: cystatin homology <CYS>  
 F:30-130/Domain: propeptide #status predicted <PRO>  
 F:131-148/Product: protegrin 5 #status predicted <MAT>  
 F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from following g)

Query Match 57.3%; Score 51; DB 2; Length 149;  
 Best Local Similarity 61.1%; Pred. No. 0.21;  
 Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18  
 ||||| | || : |||  
 Db 131 RGGRLCYCRRRRCVGVGR 148

## RESULT 3

JN0900  
 protegrin 2 precursor - pig  
 N:Alternate names: cathelin-like protein precursor; neutrophil peptide 3  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999  
 C:Accession: JN0900; S36822; S34586  
 R:Storici, P.; Zanetti, M.  
 Biochem. Biophys. Res. Commun. 196, 1363-1368, 1993  
 A:Title: A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide with a cat  
 A:Reference number: JN0900; MUID:94071898  
 A:Accession: JN0900  
 A:Molecule type: mRNA  
 A:Residues: 1-147 <STO>  
 A:Cross-references: GB:L24745; NID:g431435; PIDN:AAA31061.1; PID:g431436  
 R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Egorov,  
 FEBS Lett. 330, 339-342, 1993  
 A:Title: Primary structure of three cationic peptides from porcine neutrophils. Sequence  
 A:Reference number: S36820; MUID:93387466  
 A:Accession: S36822  
 A:Molecule type: protein  
 A:Residues: 131-146 <MIR>  
 R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.; Sh  
 FEBS Lett. 327, 231-236, 1993  
 A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of corticost  
 A:Reference number: S34585; MUID:93327946  
 A:Accession: S34586  
 A:Molecule type: protein  
 A:Residues: 131-146 <KOK>  
 C:Comment: This protein plays a role in the nonoxidative antimicrobial defense mechanism

C:Superfamily: cathelin; cystatin homology  
 C:Keywords: amidated carboxyl end; antibacterial; neutrophil  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 F:22-129/Domain: cystatin homology <CYS>  
 F:131-146/Product: protegrin 2 #status experimental <MAT>  
 F:146/Modified site: amidated carboxyl end (Val) (amide in mature form from followin

Query Match 53.9%; Score 48; DB 2; Length 147;  
 Best Local Similarity 64.7%; Pred. No. 0.66;  
 Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 17  
 ||||| | || : |||  
 Db 131 RGGRLCYCRRRRCVGVGR 147

## RESULT 4

A53895  
 protegrin 3 precursor - pig  
 N:Alternate names: neutrophil peptide 2  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 16-Jul-1999  
 C:Accession: S66285; A53895; S34587; S36821; S57608  
 R:Zhao, C.; Ganz, T.; Lehrer, R.I.  
 FEBS Lett. 368, 197-202, 1995  
 A:Title: The structure of porcine protegrin genes.  
 A:Reference number: S66283; MUID:95354835  
 A:Accession: S66285  
 A:Molecule type: DNA  
 A:Residues: 1-149 <ZH3>  
 A:Cross-references: EMBL:X84095; NID:g887644; PIDN:CAA58891.1; PID:g887645  
 R:Zhao, C.; Liu, L.; Lehrer, R.I.  
 FEBS Lett. 346, 285-288, 1994  
 A:Title: Identification of a new member of the protegrin family by cDNA cloning.  
 A:Reference number: S45712; MUID:94283613  
 A:Accession: A53895  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-149 <ZHA>  
 A:Cross-references: GB:X83267; NID:g603037; PIDN:CAA58240.1; PID:g603038  
 R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.  
 FEBS Lett. 327, 231-236, 1993  
 A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of corti  
 A:Reference number: S34585; MUID:93327946  
 A:Accession: S34587  
 A:Molecule type: protein  
 A:Residues: 131-148 <KOK>  
 R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Ego  
 FEBS Lett. 330, 339-342, 1993  
 A:Title: Primary structure of three cationic peptides from porcine neutrophils. Sequ  
 A:Reference number: S36820; MUID:93387466  
 A:Accession: S36821  
 A:Molecule type: protein  
 A:Residues: 131-148 <MIR>  
 C:Genetics:  
 A:Gene: NPG3  
 A:Introns: 66/3; 102/3; 126/3  
 C:Superfamily: cathelin; cystatin homology  
 C:Keywords: amidated carboxyl end; antibacterial; neutrophil  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 F:22-129/Domain: cystatin homology <CYS>  
 F:30-130/Domain: propeptide #status predicted <PRO>  
 F:131-148/Product: protegrin 3 #status experimental <MAT>  
 F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followin

Query Match 50.6%; Score 45; DB 2; Length 149;  
 Best Local Similarity 55.6%; Pred. No. 2.1;  
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18  
 ||||| | || : |||

Db 131 RGGGLCYCRRRCVCVGR 148

RESULT 5

C83023

probable oxidoreductase PA4986 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C83023

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brody, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, S.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A:Reference number: A82950; MUID:20437337

A:Accession: C83023

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-648 <STO>

A:Cross-references: GB:AE004911; GB:AE004091; NID:99951264; PIDN:AAG08371.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4986

C:Superfamily: Methylotrophus W3A1 trimethylamine dehydrogenase

Query Match 50.6%; Score 45; DB 2; Length 648;

Best Local Similarity 50.0%; Pred. No. 8.4;

Matches 11; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

Qy 1 RGGRL----AYRLRLFAIRVGR 18

||||| :||| :|||

Db 487 RGGELQVVDQWGLRGEVRVGR 508

RESULT 6

H69088

cell surface glycoprotein (s-layer protein) related protein - Methanobacterium thermoautotrophicum

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999

C:Accession: H69088

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N. K.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: function

A:Reference number: A69000; MUID:98037514

A:Accession: H69088

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1408 <MTH>

A:Cross-references: GB:AE000911; GB:AE000666; NID:g2622623; PIDN:AAB85988.1; PID:g2622623

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1513

A:Start codon: TTG

Query Match 49.4%; Score 44; DB 2; Length 1408;

Best Local Similarity 41.2%; Pred. No. 26;

Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GGRLAYRLRLFAIRVGR 18

||||| :||| :|||

Db 1318 GAKLYRFTFKLPGR 1334

RESULT 7

C69439

sugar fermentation stimulation protein (sfsA) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000

C:Accession: C69439

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dooley, D.P.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Pyrococcus furiosus

A:Reference number: A69250; MUID:98049343

A:Accession: C69439

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-219 <KLE>

A:Cross-references: GB:AE000997; GB:AE000782; NID:g2689320; PIDN:AAB89729.1; PID:g2689320

Query Match 47.2%; Score 42; DB 2; Length 219;

Best Local Similarity 64.3%; Pred. No. 9.9;

Matches 9; Conservative 3; Mismatches 0; Indels 2; Gaps 1;

Qy 1 RGGRLAYRLRLFAI 14

||||| :||| :|||

Db 58 RGGKLSYRL--FAV 69

RESULT 8

D75587

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: D75587

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R. M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C. S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1

A:Reference number: A75250; MUID:20036896

A:Accession: D75587

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-236 <WHI>

A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12511.1; PID:g6460670

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRA0327

A:Map position: 2

Query Match 47.2%; Score 42; DB 2; Length 236;

Best Local Similarity 62.5%; Pred. No. 11;

Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGRLAYRLRLFAIRVG 17

||||| :||| :|||

Db 73 GGELARQLLRWAARDG 88

RESULT 9

D83231

hypothetical protein PA3305 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: D83231

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brody, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A:Reference number: A82950; MUID:20437337

A:Accession: D83231

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-664 <STO>

A:Cross-references: GB:AE004753; GB:AE004091; NID:g9949433; PIDN:AAG06693.1; GSPDB:G

A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3305

Query Match 47.2%; Score 42; DB 2; Length 664;  
Best Local Similarity 75.0%; Pred. No. 28;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGRLAYRLRLRFA 13  
|||:| |||||  
Db 536 GGRMADRLRLRLA 547

RESULT 10  
A83488  
hypothetical protein PA1270 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83488  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: A83488  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-679 <STO>  
A:Cross-references: GB:AF004556; GB:AF004091; NID:99947194; PIDN:AAG04659.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1270

Query Match 47.2%; Score 42; DB 2; Length 679;  
Best Local Similarity 58.8%; Pred. No. 29;  
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGRLAYRLRLRFAIRVGR 18  
|||:| |||||  
Db 539 GGRTADRLRLRQAQYDR 555

RESULT 11  
T35824  
probable oxidoreductase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C:Accession: T35824  
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1999  
A:Reference number: Z21589  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-447 <MUR>  
A:Cross-references: EMBL:AL035569; PIDN:CAB37584.1; GSPDB:GN00070; SCOEDB:SC8D9.20c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC8D9.20c  
C:Superfamily: hypothetical protein b0837

Query Match 46.1%; Score 41; DB 2; Length 447;  
Best Local Similarity 66.7%; Pred. No. 29;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGRLAYRLRLRFAIR 15  
||:| |||||  
Db 16 RGRLAYRGSRFAAR 30

RESULT 12  
E70887

probable fadD32 protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 03-Nov-2000  
C:Accession: E70887

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen  
A:Reference number: A70500; MUID:98295987  
A:Accession: E70887  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-637 <COL>  
A:Cross-references: GB:AL022076; GB:AL123456; NID:93256026; PIDN:CAA17865.1; PID:929  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: fadD32  
C:Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology  
F:81-614/Domain: acetate--CoA ligase homology <ACL>

Query Match 46.1%; Score 41; DB 2; Length 637;  
Best Local Similarity 61.5%; Pred. No. 40;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRLRFA 13  
||:| |||||  
Db 40 RGDKLAYRFLDFS 52

RESULT 13  
QQIHC

hypothetical protein (gene N internal ORF) - bovine coronavirus (strain Mebus)  
N:Alternate names: IORF protein  
C:Species: bovine coronavirus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 16-Jul-1999  
C:Accession: C26347  
R:Lapps, W.; Hogue, B.G.; Brian, D.A.  
Virology 157, 47-57, 1987  
A:Title: Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein  
A:Reference number: A94357; MUID:87151119  
A:Molecule type: genomic RNA  
A:Residues: 1-207 <LAP>  
A:Cross-references: GB:M16620; NID:932354; PIDN:AAA66398.1; PID:9807593  
C:Superfamily: coronavirus gene N internal ORF

Query Match 44.9%; Score 40; DB 1; Length 207;  
Best Local Similarity 53.8%; Pred. No. 21;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 RLAYRLRLRFAIRV 16  
||:| |||||  
Db 153 RLGFRLAYRLSLRV 165

RESULT 14  
JQ1174

hypothetical protein (gene N internal ORF) - turkey coronavirus  
N:Alternate names: IORF protein  
C:Species: turkey coronavirus  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
C:Accession: JQ1174  
R:Verbeek, A.; Tijssen, P.  
J. Gen. Virol. 72, 1659-1666, 1991  
A:Title: Sequence analysis of the turkey enteric coronavirus nucleocapsid and membran  
A:Reference number: JQ1172; MUID:91311418

A:Accession: JQ1174  
A:Molecule type: genomic RNA  
A:Residues: 1-207 <VER>  
A:Experimental source: strain Minnesota  
C:Comment: The gene encoding this protein is located within the nucleocapsid (N) gene.  
C:Superfamily: coronavirus gene N internal ORF

Query Match 44.9%; Score 40; DB 1; Length 207;  
Best Local Similarity 53.8%; Pred. No. 21;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

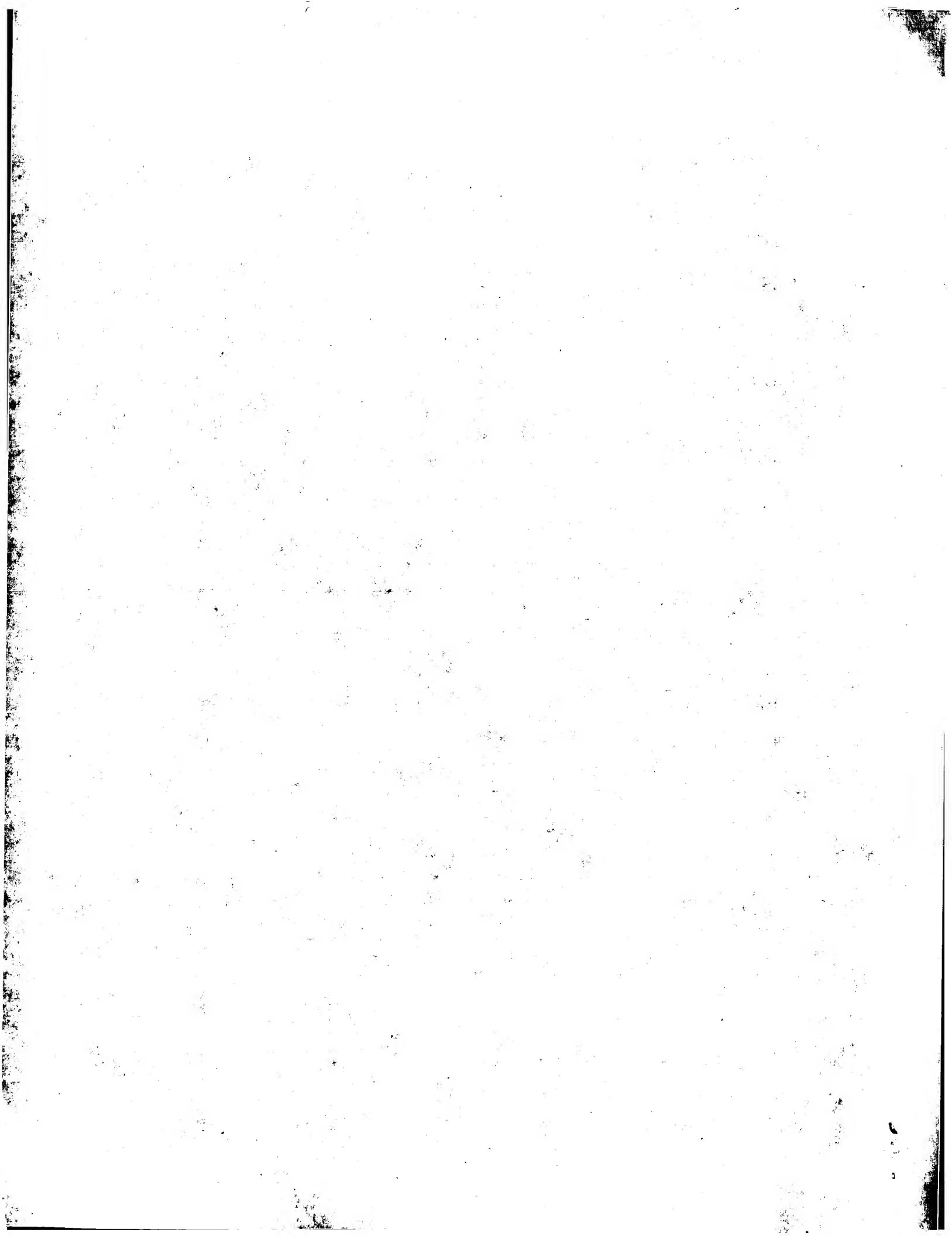
QY 4 RLAYRLLRFAIRV 16  
|| :|| |::||  
Db 153 RLGFRLARYSLRV 165

RESULT 15  
S06869  
hypothetical protein (gene N internal ORF) - bovine coronavirus (strain F15)  
N:Alternate names: IORF protein  
C:Species: bovine coronavirus  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 16-Jul-1999  
C:Accession: S06869  
R:Cruciere, C.; Laporte, J.  
Ann. Inst. Pasteur Virol. 139, 123-138, 1988  
A:Title: Sequence and analysis of bovine enteritic coronavirus (F15) genome. I. - Sequen  
A:Reference number: S06399; WUID:89087718  
A:Accession: S06869  
A:Molecule type: genomic RNA  
A:Residues: 1-207 <CRU>  
A:Cross-references: EMBL:M36656; NID:g210700; PIDN:AAA42759.1; PID:g210702  
A:Note: the source is designated as bovine enteritic coronavirus  
C:Comment: The gene encoding this protein is located within the nucleocapsid (N) gene.  
C:Superfamily: coronavirus gene N internal ORF

Query Match 44.9%; Score 40; DB 1; Length 207;  
Best Local Similarity 53.8%; Pred. No. 21;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 RLAYRLLRFAIRV 16  
|| :|| |::||  
Db 153 RLGFRLARYSLRV 165

Search completed: February 12, 2002, 12:34:37  
Job time: 555 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:22 ; Search time 106.12 Seconds  
(without alignments)  
3.817 Million cell updates/sec

Title: US-09-485-571-19

Perfect score: 89

Sequence: 1 RGGRLAYRLRFAIRVGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	65.2	18	2	US-08-752-852A-230
2	56	62.9	18	1	US-08-499-523-37
3	56	62.9	18	1	US-08-499-523-63
4	56	62.9	18	4	US-09-128-345-63
5	56	62.9	18	4	US-09-128-345-67
6	55	61.8	18	1	US-08-499-523-54
7	55	61.8	18	1	US-08-499-523-59
8	55	61.8	18	4	US-09-128-345-54
9	55	61.8	18	4	US-09-128-345-59
10	54	60.7	18	1	US-08-499-523-53
11	54	60.7	18	1	US-08-499-523-58
12	54	60.7	18	2	US-08-752-852A-233
13	54	60.7	18	4	US-09-128-345-53
14	54	60.7	18	4	US-09-128-345-58
15	53	59.6	18	1	US-08-499-523-57
16	53	59.6	18	1	US-08-499-523-62
17	53	59.6	18	2	US-08-752-852A-128
18	53	59.6	18	4	US-09-128-345-57
19	53	59.6	18	4	US-09-128-345-62
20	52	58.4	18	1	US-08-095-769A-1
21	52	58.4	18	1	US-08-182-483A-2
22	52	58.4	18	1	US-08-182-483A-17
23	52	58.4	18	1	US-08-243-879A-1
24	52	58.4	18	1	US-08-243-879A-16
25	52	58.4	18	1	US-08-499-523-11
26	52	58.4	18	1	US-08-499-523-16
27	52	58.4	18	1	US-08-499-523-33

RESULT 1  
US-08-752-852A-230  
; Sequence 230, Application US/08752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrer, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 230:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-752-852A-230

ALIGNMENTS

28	52	58.4	18	1	US-08-499-523-37	Sequence 37, Appl
29	52	58.4	18	2	US-08-752-852A-1	Sequence 1, Appl1
30	52	58.4	18	2	US-08-752-852A-123	Sequence 123, App
31	52	58.4	18	3	US-08-752-853-1	Sequence 1, Appl1
32	52	58.4	18	3	US-08-752-853-2	Sequence 2, Appl1
33	52	58.4	18	3	US-08-984-294-1	Sequence 1, Appl1
34	52	58.4	18	4	US-09-128-345-11	Sequence 11, Appl
35	52	58.4	18	4	US-09-128-345-16	Sequence 16, Appl
36	52	58.4	18	4	US-09-128-345-33	Sequence 33, Appl
37	52	58.4	18	4	US-09-128-345-37	Sequence 37, Appl
38	52	58.4	149	1	US-08-243-879A-36	Sequence 36, Appl
39	52	58.4	149	1	US-08-499-523-2	Sequence 2, Appl1
40	52	58.4	149	1	US-08-499-523-10	Sequence 10, Appl
41	52	58.4	149	4	US-09-128-345-2	Sequence 2, Appl1
42	52	58.4	149	4	US-09-128-345-10	Sequence 10, Appl
43	51	57.3	18	1	US-08-182-483A-26	Sequence 26, Appl
44	51	57.3	18	1	US-08-243-879A-25	Sequence 25, Appl
45	51	57.3	18	1	US-08-499-523-20	Sequence 20, Appl

Query Match 65.2%; Score 58; DB 2; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.0015; 5; Indels  
Matches 12; Conservative 1; Mismatches 5; Gaps 0;

Qy 1 RGGRLAYRLLRFAIRVGR 18  
||||| | ||| : |||  
Db 1 RGGRLCYARRRFAVCVGR 18

## RESULT 2

US-08-499-523-63  
; Sequence 63, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 8, 13, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-63

Query Match 62.9%; Score 56; DB 1; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.0033;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RGGRLAYRLLRFAIRVGR 18  
||||| | ||| : |||  
Db 1 RGGRLXYXRRRFXVXVGR 18

## RESULT 3

US-08-499-523-67  
; Sequence 67, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.

; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 67:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 8, 13, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-67

Query Match 62.9%; Score 56; DB 1; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.0033;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RGGRLAYRLLRFAIRVGR 18  
||||| | ||| : |||  
Db 1 RGGRLXYXRRRFXVXVGR 18

## RESULT 4

US-09-128-345-63  
; Sequence 63, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.

; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A. 742
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-63
```

```
Query Match 62.9%; Score 56; DB 4; Length 18;
Best Local Similarity 61.1%; Pred. No. 0.0033;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 RGGRLAYRLRLFAIRVGR 18
   ||||| | || |||
Db 1 RGGRLXYRRRFXVXVGR 18
```

## RESULT 5

```
US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-67
```

```
Query Match 62.9%; Score 56; DB 4; Length 18;
Best Local Similarity 61.1%; Pred. No. 0.0033;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 RGGRLAYRLRLFAIRVGR 18
   ||||| | || |||
Db 1 RGGRLXYRRRFXVXVGR 18
```

## RESULT 6

```
US-08-499-523-54
; Sequence 54, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 8..13
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-08-499-523-54
```

```
Query Match 61.8%; Score 55; DB 1; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.0048;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```

QY 1 RGGRLAYLLRFAIRVGR 18
    ||||| | | | | |
Db 1 RGGRLXYCRRRFCIXVGR 18

RESULT 7
US-08-499-523-59
; Sequence 59, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MORASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 6..15
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(8, 13)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-08-499-523-59

Query Match 61.8%; Score 55; DB 1; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.0048;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYLLRFAIRVGR 18
    ||||| | | | | |
Db 1 RGGRLXYCRRRFCIXVGR 18

RESULT 8
US-09-128-345-54
; Sequence 54, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 8..13
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-54

Query Match 61.8%; Score 55; DB 4; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.0048;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYLLRFAIRVGR 18
    ||||| | | | | |
Db 1 RGGRLXYCRRRFCIXVGR 18

RESULT 9
US-09-128-345-59
; Sequence 59, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 6..15  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(8, 13)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-59

Query Match 61.8%; Score 55; DB 4; Length 18;  
Best Local Similarity 66.7%; Pred. No. 0.0048;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18  
| | | | | | | | | | | | | | | | | | | |  
Db 1 RGGRLCYRRRFXICVGR 18

RESULT 10  
US-08-499-523-53  
Sequence 53, Application US/08499523  
Patent No. 5804558  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KORYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,523  
FILING DATE: 07-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 8..13  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-53

Query Match 60.7%; Score 54; DB 1; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.007;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18  
| | | | | | | | | | | | | | | | | | | |  
Db 1 RGGRLCYRRRFXICVGR 18

RESULT 11  
US-08-499-523-58  
Sequence 58, Application US/08499523  
Patent No. 5804558  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KORYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,523  
FILING DATE: 07-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 6..15  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(8, 13)  
OTHER INFORMATION: /note= "X is a hydrophobic, a

OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-58

Query Match 60.7%; Score 54; DB 1; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.007;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRLFAIRVGR 18  
||||| | | | |  
Db 1 RGGRLCYRRRRCVXVGR 18

## RESULT 12

US-08-752-852A-233  
; Sequence 233, Application US/08752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrer, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 233:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-752-852A-233

Query Match 60.7%; Score 54; DB 2; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.007;  
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRLFAIRVGR 18  
||| | | | | | | |  
Db 1 RGGGLCYRLPKFRVCVGR 18

## RESULT 13

-4-

US-09-128-345-53  
; Sequence 53, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-0054-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Disulfide-bond  
; LOCATION: 8..13  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-53

Query Match 60.7%; Score 54; DB 4; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.007;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRLFAIRVGR 18  
||||| | | | |  
Db 1 RGGRLCYRRRRCVXVGR 18

## RESULT 14

US-09-128-345-58  
; Sequence 58, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York

STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
NAME/KEY: Disulfide-bond  
LOCATION: 6..15  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(8, 13)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
US-09-128-345-58  
OTHER INFORMATION: small, or a large polar amino acid"

Query Match 60.7%; Score 54; DB 4; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.007;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RGGRLAYLLRFAIRVGR 18  
||||| | | | : |||  
Db 1 RGGRLCYXRRRFVCVGR 18

RESULT 15  
US-08-499-523-57  
Sequence 57, Application US/08499523  
Patent No. 5804558  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,523  
FILING DATE: 07-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 8..13  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
US-08-499-523-57  
OTHER INFORMATION: small, or a large polar amino acid"

Query Match 59.6%; Score 53; DB 1; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.01;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RGGRLAYLLRFAIRVGR 18  
||||| | | | : |||  
Db 1 RGGRLCYXRRRFVCVGR 18

Search completed: February 12, 2002, 12:32:22  
Job time: 450 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:30 ; Search time 242.57 Seconds  
(without alignments)  
5.497 Million cell updates/sec

Title: US-09-485-571-19

Perfect score: 89

Sequence: 1 RGRRLAYLLRFAIRVGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	18	20 AAW99406	Protegrin derivati
2	59	66.3	18	18 AAW18151	Cationic, antimicr
3	59	66.3	18	18 AAW18152	Cationic, antimicr
4	58	65.2	18	18 AAW36429	Antimicrobial prot
5	58	65.2	18	18 AAW09084	Cationic, antimicr
6	58	65.2	18	18 AAW09085	Cationic, antimicr
7	58	65.2	18	20 AAW99403	Protegrin derivati
8	58	65.2	18	21 AAW93616	Peptide which may
9	57	64.0	18	18 AAW18153	Cationic, antimicr
10	57	64.0	18	21 AAY93669	Peptide which may
11	54	60.7	18	18 AAW36432	Antimicrobial prot

12	54	60.7	18	20 AAW99408	Protegrin derivati
13	53	59.6	18	18 AAW18147	Cationic, antimicr
14	53	59.6	18	18 AAW18148	Cationic, antimicr
15	53	59.6	18	18 AAW18149	Cationic, antimicr
16	52	58.4	18	16 AAR78751	Protegrin PG-1. S
17	52	58.4	18	16 AAR78765	Protegrin peptide
18	52	58.4	18	18 AAW36322	Antimicrobial prot
19	52	58.4	18	18 AAW36208	Antimicrobial prot
20	52	58.4	18	18 AAW36353	Antimicrobial prot
21	52	58.4	18	18 AAW35578	Antimicrobial pept
22	52	58.4	18	18 AAW18144	Cationic, antimicr
23	52	58.4	18	18 AAW18130	Cationic, antimicr
24	52	58.4	18	18 AAW09073	Cationic, antimicr
25	52	58.4	18	19 AAW29556	Porcine protegrin
26	52	58.4	18	19 AAW66458	Cationic peptide p
27	52	58.4	18	20 AAY22018	Protegrin peptide
28	52	58.4	18	21 AAY93170	Protegrin peptide
29	52	58.4	18	21 AAY93608	Protegrin peptide
30	52	58.4	18	21 AAY81680	Protegrin peptide
31	52	58.4	18	21 AAY91757	Cationic peptide p
32	52	58.4	18	22 AAB91843	Antimicrobial pept
33	52	58.4	18	22 AAB35050	Porcine protegrin-
34	52	58.4	149	18 AAW25081	Antimicrobial comp
35	52	58.4	149	18 AAW09087	Antimicrobial prot
36	51	57.3	18	16 AAR78774	Protegrin peptide
37	51	57.3	18	18 AAW36329	Antimicrobial prot
38	51	57.3	18	18 AAW35581	Antimicrobial pept
39	51	57.3	18	18 AAW18150	Cationic, antimicr
40	51	57.3	18	18 AAW18137	Cationic, antimicr
41	51	57.3	18	18 AAW09077	Cationic, antimicr
42	51	57.3	18	19 AAW29560	Porcine protegrin
43	51	57.3	18	21 AAB11030	Porcine protegrin
44	51	57.3	18	21 AAY93174	Protegrin peptide
45	51	57.3	18	21 AAY93612	Protegrin peptide

#### ALIGNMENTS

#### RESULT 1

AAW99406  
ID AAW99406 standard; peptide; 18 AA.

XX AAW99406;

XX AC

XX 08-JUN-1999 (first entry)

XX Protegrin derivative peptide SM2187.

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

XX Synthetic.

XX WO9907728-A2.

XX 18-FEB-1999.

XX 06-AUG-1998; 98WO-FR01757.

XX 12-AUG-1997; 97FR-0010297.

XX (SYNT-) SYNT:EM SA.

XX Calas E, Chavanieu A, Grassy G, Kaczorek M;

XX WPI; 1999-190034/16.

XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
PT as carriers to deliver active agents into cells  
XX



CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. This peptide is in snake form where all the  
 CC cysteine residues are replaced by a hydrophobic, small or large polar  
 CC amino acid (e.g. alanine in this case). Peptides of this formula are  
 CC designated protegrins and are useful as anti-bacterial, anti-viral and  
 CC anti-fungal agents in plants and animals. The protegrins confer  
 CC resistance to microbial or viral infection in plants by preventing the  
 CC growth of a virus or microbe and inactivate the endotoxin of gram-  
 CC negative bacteria. The protegrins are particularly useful for the  
 CC treatment of sexually transmitted disease caused by microorganisms e.g.  
 CC *Candida albicans*, HIV-1, *Chlamydia trachomatis*, *Treponema pallidum* and  
 CC *Neisseria gonorrhoeae*. They can also be used in eye care solutions and  
 CC as preservatives for food. The protegrins are more effective under  
 CC physiological conditions (e.g. in the presence of serum) than certain  
 CC antibiotics and are non-toxic to the cells of higher organisms.

XX Sequence 18 AA;

Query Match 66.3%; Score 59; DB 18; Length 18;

Best Local Similarity 66.7%; Pred. No. 0.0032;

Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRFAIRVGR 18

Db 1 rggrlayrrrravavgr 18

RESULT 4

ID AAW36429 standard; peptide; 18 AA.

AC AAW36429;

DT 13-FEB-1998 (first entry)

DE Antimicrobial protegrin peptide (229).

XX Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW *Helicobacter pylori*; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant *Enterococcus*; pathogen; multi-drug resistance;  
 KW penicillin resistant *Streptococcus pneumoniae*; pig; porcine;  
 KW methicillin resistant *Staphylococcus aureus*; systemic candidiasis.

XX Synthetic.

OS Sus scrofa.

PN WO9718826-A1.

XX 29-MAY-1997.

XX 22-NOV-1996; 96WO-US18544.

XX 21-NOV-1996; 96US-0752852.

PR 22-NOV-1995; 95US-0562346.

PR 17-MAY-1996; 96US-0649811.

PR 01-AUG-1996; 96US-0690921.

XX (INTR-) INTRABIOTICS PHARM INC.

(REGC ) UNIV CALIFORNIA.

Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;

WPI; 1997-297871/27.

XX New antimicrobial protegrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)

PS Claim 23; Page 110; 130pp; English.

XX The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoan and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC *Helicobacter pylori*, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC *Enterococcus faecium* or faecalis, penicillin resistant  
 CC *Streptococcus pneumoniae* and methicillin resistant *Staphylococcus*  
 CC *aureus* (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.

XX Sequence 18 AA;

Query Match 65.2%; Score 58; DB 18; Length 18;

Best Local Similarity 66.7%; Pred. No. 0.0046;

Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRFAIRVGR 18

Db 1 rggrlayrrrravavgr 18

RESULT 5

AAW09084

ID AAW09084 standard; peptide; 18 AA.

AC AAW09084;

DT 11-AUG-1997 (first entry)

XX Cationic, antimicrobial, virus-neutralising protegrin IB-288.

XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;

KW *Candida albicans*; gram-negative bacteria; STD;

KW sexually transmitted disease; HIV-1; *Chlamydia trachomatis*;

KW *Treponema pallidum*; *Neisseria gonorrhoeae*; eye care; preservative;  
 KW food.

XX Synthetic.

OS Key Location/Qualifiers

XX Modified-site 1

FT /note= "Acylated"

FT Disulfide-bond 6..15

FT /note= "results in bullet form peptide"

XX WO9637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

```

XX 07-JUL-1995; 95US-0499523.
XX 26-MAY-1995; 95US-0451832.
XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.
XX
XX Harwig SSL, Kokryakov VN, Lehrer RI;
XX WPI; 1997-033984/03.
XX
XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -
XX useful for the treatment of microbial infection, as food
XX preservatives and in eye care solutions
XX
XX Claim 6; Page 65; 106pp; English.
XX
XX The present sequence is a specifically claimed example of a peptide,
XX recombinantly produced, corresponding to the generic formula:
XX A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)
XX where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a
XX basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;
XX A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino
XX acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino
XX acid or proline; A17 may be absent or a basic, neutral/polar,
XX hydrophobic or small amino acid; and A18 may be absent or a basic,
XX neutral/polar, hydrophobic or small amino acid. This has a charge of at
XX least +3 and its N-terminal acylated and/or C-terminal amidated or
XX cysteine bridge. Peptides of this formula are designated protegrins and
XX are useful as anti-bacterial, anti-viral and anti-fungal agents in
XX plants and animals. The protegrins confer resistance to microbial or
XX viral infection in plants by preventing the growth of a virus or microbe
XX and inactivate the endotoxin of gram-negative bacteria. The protegrins
XX are particularly useful for the treatment of sexually transmitted
XX disease caused by microorganisms e.g. Candida albicans, HIV-1, Chlamydia
XX trachomatis, Treponema pallidum and Neisseria gonorrhoeae. They can also
XX be used in eye care solutions and as preservatives for food. The
XX protegrins are more effective under physiological conditions (e.g. in
XX the presence of serum) than certain antibiotics and are non-toxic to the
XX cells of higher organisms.
XX
XX Sequence 18 AA;

Query Match 65.2%; Score 58; DB 18; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.0046;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18
Db 1 rggrlcayrrrfavcvgvr 18
||||| | |||: |||

RESULT 6
AAW09085
ID AAW09085 standard; peptide; 18 AA.
XX
XX AAW09085;
XX
XX 11-AUG-1997 (first entry)
XX
XX Cationic, antimicrobial, virus-neutralising protegrin IB-289.
XX
XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;
XX Candida albicans; gram-negative bacteria; STD;
XX sexually transmitted disease; HIV-1; Chlamydia trachomatis;
XX Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;
XX food.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1
XX
XX Protegrin derivative peptide SM1738.

```

```

FT Disulfide-bond /note= "Acylated"
FT 6..15
FT /note= "results in bullet form peptide"
FT 18
FT Modified-site /note= "Amidated"
FT 18
XX
XX WO9637508-A1.
XX
XX 28-NOV-1996.
XX
XX 24-MAY-1996; 96WO-US07594.
XX
XX 07-JUL-1995; 95US-0499523.
XX 26-MAY-1995; 95US-0451832.
XX
XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.
XX
XX Harwig SSL, Kokryakov VN, Lehrer RI;
XX WPI; 1997-033984/03.
XX
XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -
XX useful for the treatment of microbial infection, as food
XX preservatives and in eye care solutions
XX
XX Claim 6; Page 65; 106pp; English.
XX
XX The present sequence is a specifically claimed example of a peptide,
XX recombinantly produced, corresponding to the generic formula:
XX A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)
XX where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a
XX basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;
XX A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino
XX acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino
XX acid or proline; A17 may be absent or a basic, neutral/polar,
XX hydrophobic or small amino acid; and A18 may be absent or a basic,
XX neutral/polar, hydrophobic or small amino acid. This has a charge of at
XX least +3 and its N-terminal acylated and/or C-terminal amidated or
XX cysteine bridge. Peptides of this formula are designated protegrins and
XX are useful as anti-bacterial, anti-viral and anti-fungal agents in
XX plants and animals. The protegrins confer resistance to microbial or
XX viral infection in plants by preventing the growth of a virus or microbe
XX and inactivate the endotoxin of gram-negative bacteria. The protegrins
XX are particularly useful for the treatment of sexually transmitted
XX disease caused by microorganisms e.g. Candida albicans, HIV-1, Chlamydia
XX trachomatis, Treponema pallidum and Neisseria gonorrhoeae. They can also
XX be used in eye care solutions and as preservatives for food. The
XX protegrins are more effective under physiological conditions (e.g. in
XX the presence of serum) than certain antibiotics and are non-toxic to the
XX cells of higher organisms.
XX
XX Sequence 18 AA;

Query Match 65.2%; Score 58; DB 18; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.0046;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18
Db 1 rggrlcayrrrfavcvgvr 18
||||| | |||: |||

RESULT 7
AAW99403
ID AAW99403 standard; peptide; 18 AA.
XX
XX AAW99403;
XX
XX 08-JUN-1999 (first entry)
XX
XX Protegrin derivative peptide SM1738.

```

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.  
XX Synthetic.  
OS  
XX WO9907728-A2.  
PN  
XX  
PD 18-FEB-1999.  
XX  
PF 06-AUG-1998; 98WO-FR01757.  
XX  
PR 12-AUG-1997; 97FR-0010297.  
XX  
PA (SYNT-) SYNT:EM SA.  
XX Calas B, Chavanieu A, Grassy G, Kaczorek M;  
XX WPI; 1999-190034/16.  
DR  
XX Derivatives of antibiotic peptides lacking disulfide bridges - used  
XX as carriers to deliver active agents into cells  
PT  
XX  
PS Claim 7; Page 28; 37pp; French.  
XX  
XX This peptide represents a linear derivative of the protegrin family of  
CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
CC antibiotic family which contain a beta-sheet secondary structure linked  
CC by disulphide bridges. The new derivatives are linear and lack the  
CC disulphide bridge. The novel derivatives are used to deliver active  
CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
XX including crossing the blood-brain barrier.  
XX  
SQ Sequence 18 AA;

Query Match 65.2%; Score 58; DB 20; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.0046;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGGRLAYRLRLFAIRVGR 18  
||||| | | | | |  
Db 1 rggrrlsyrrrrfsvsvgr 18

RESULT 8  
AAV93616  
ID AAV93616 standard; peptide; 18 AA.  
XX  
AC AAV93616;

XX 25-SEP-2000 (first entry)  
XX  
XX Peptide which may be linked to anticancer agents.  
DE  
XX  
XX Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
KW cancer.  
KW  
XX Unidentified.  
OS  
XX WO200032237-A1.  
PN  
XX 08-JUN-2000.  
PD  
XX 26-NOV-1999; 99WO-FR02939.  
PF  
XX 30-NOV-1998; 98FR-0015073.  
PR

XX  
PA (SYNT-) SYNT:EM SA.  
XX Tensamani J, Kaczorek M, Colin De Verdiere A;  
PI  
XX WPI; 2000-412166/35.  
DR  
XX  
XX New composition useful for cancer treatment and prevention, contains  
PT anticancer agent and peptide vector that transports agent into cells  
PT  
XX Disclosure; Page 8; 34pp; French.  
PS  
XX  
XX The specification describes a pharmaceutical composition, which  
CC comprises at least one anticancer agent associated with at least one  
CC peptide that can transport it into cancer cells and which inhibits  
CC development of resistance to the anticancer agent. By using the  
CC peptide as a vector for delivery of the anticancer agent, mechanisms  
CC that cause cancer cells to become resistant to the agent, particularly  
CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
CC produced by chemical synthesis, can be coupled easily to the agent,  
CC cross mammalian cell membranes rapidly by a passive mechanism (no  
CC receptors required), and are non-toxic and non-lytic. The compositions  
CC are used to treat cancer. The present sequence represents a peptide  
CC which may be linked to the anticancer agents of the invention.  
XX  
SQ Sequence 18 AA;

Query Match 65.2%; Score 58; DB 21; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.0046;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGGRLAYRLRLFAIRVGR 18  
||||| | | | | |  
Db 1 rggrrlsyrrrrfsvsvgr 18

RESULT 9  
AAW18153  
ID AAW18153 standard; peptide; 18 AA.  
XX  
AC AAW18153;

XX 11-AUG-1997 (first entry)  
XX  
XX Cationic, antimicrobial, virus-neutralising protegrin PC-57.  
DE  
XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
KW Candida albicans; gram-negative bacteria; STD;  
KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;  
KW food.  
KW  
XX Synthetic.  
OS  
XX WO9637508-A1.  
PN  
XX 28-NOV-1996.  
XX  
XX 24-MAY-1996; 96WO-US07594.  
PF  
XX 07-JUL-1995; 95US-0499523.  
PR  
XX 26-MAY-1995; 95US-0451832.  
XX  
XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.  
PA  
XX Harwig SSL, Kokryakov VN, Lehrer RI;  
XX WPI; 1997-033984/03.  
XX  
XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
PT useful for the treatment of microbial infection, as food  
PT preservatives and in eye care solutions

XX Claim 6; Page 64; 106pp; English.  
 XX  
 CC The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. This peptide is in snake form where all the  
 CC cysteine residues are replaced by a hydrophobic, small or large polar  
 CC amino acid (e.g. alanine in this case). Peptides of this formula are  
 CC designated protegrins and are useful as anti-bacterial, anti-viral and  
 CC anti-fungal agents in plants and animals. The protegrins confer  
 CC resistance to microbial or viral infection in plants by preventing the  
 CC growth of a virus or microbe and inactivate the endotoxin of gram-  
 CC negative bacteria. The protegrins are particularly useful for the  
 CC treatment of sexually transmitted disease caused by microorganisms e.g.  
 CC Candida albicans, HIV-1, Chlamydia trachomatis, Treponema pallidum and  
 CC Neisseria gonorrhoeae. They can also be used in eye care solutions and  
 CC as preservatives for food. The protegrins are more effective under  
 CC physiological conditions (e.g. in the presence of serum) than certain  
 CC antibiotics and are non-toxic to the cells of higher organisms.  
 XX  
 SQ Sequence 18 AA;

Query Match 64.0%; Score 57; DB 18; Length 18;  
 Best Local Similarity 66.7%; Pred. No. 0.0067;  
 Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRFAIRVGR 18  
 Db 1 rggrlayrrrfavawgr 18  
 ||||| |||

## RESULT 10

AA93669  
 ID AAY93669 standard; peptide; 18 AA.

XX AC AAY93669;

XX DT 25-SEP-2000 (first entry)

XX DE Peptide which may be linked to anticancer agents.

XX KW Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 XX cancer.

XX XX Unidentified.

XX OS WO200032237-A1.

XX PN 08-JUN-2000.

XX PD 26-NOV-1999; 99WO-FR02939.

XX PF 30-NOV-1998; 98FR-0015073.

XX PR (SYNT-) SYNT:EM SA.

XX PA Tamsamani J, Kaczorek M, Colin De Verdiere A;

XX PI WPT: 2000-412166/35.

XX DR New composition useful for cancer treatment and prevention, contains

XX anticancer agent and peptide vector that transports agent into cells

PT

XX

PS Disclosure; Page 8; 34pp; French.

XX

CC The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.

XX SQ Sequence 18 AA;

Query Match 64.0%; Score 57; DB 21; Length 18;  
 Best Local Similarity 66.7%; Pred. No. 0.0067;  
 Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLAYRLRFAIRVGR 18

Db 1 rggrlayrrrwavlgr 18  
 ||||| |:: |||

## RESULT 11

AAW36432  
 ID AAW36432 standard; peptide; 18 AA.

XX AC AAW36432;

XX DT 13-FEB-1998 (first entry)

XX DE Antimicrobial protegrin peptide (232).

XX KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biofouling; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.

XX OS Synthetic.

XX OS Sus scrofa.

XX PN WO9718826-A1.

XX PD 29-MAY-1997.

XX PF 22-NOV-1996; 96WO-US18544.

XX PR 21-NOV-1996; 96US-0752852.

XX PR 22-NOV-1995; 95US-0562346.

XX PR 17-MAY-1996; 96US-0649811.

XX PR 01-AUG-1996; 96US-0690921.

XX PA (INTR-) INTRABIOTICS PHARM INC.  
 XX (REGC ) UNIV CALIFORNIA.

XX PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;

XX WPI: 1997-297871/27.

XX New antimicrobial protegrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT

;

PT (e.g. HIV)  
 PS Claim 23; Page 111; 130pp; English.  
 XX  
 CC The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX  
 SQ Sequence 18 AA;

Query Match 60.7%; Score 54; DB 18; Length 18;  
 Best Local Similarity 61.1%; Pred. No. 0.021;  
 Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RGGRLAYLLRFAIRVGR 18  
 Db 1 rgggclcyrlpkfrvcvgr 18

## RESULT 12

AAW99408  
 ID AAW99408 standard; peptide; 18 AA.

XX AAW99408;

XX 08-JUN-1999 (first entry)

XX Protegrin derivative peptide SM2189.

XX Linear: protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX Synthetic.

XX Key Location/Qualifiers  
 FT Modified-site 9..10  
 FT /label= Nle  
 FT Modified-site 14  
 FT /label= Nva

XX W09907728-A2.

XX 18-FEB-1999.

XX 06-AUG-1998; 98WO-FR01757.

XX 12-AUG-1997; 97FR-0010297.

XX (SYNT-) SYNT:EM SA.

XX Calas B, Chavanieu A, Grassy G, Kaczorek M;

XX WPI; 1999-190034/16.

XX

PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
 XX as carriers to deliver active agents into cells  
 PS Claim 7: Page 28; 37pp; French.

XX This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX Sequence 18 AA;

Query Match 60.7%; Score 54; DB 20; Length 18;  
 Best Local Similarity 61.1%; Pred. No. 0.021;  
 Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RGGRLAYLLRFAIRVGR 18  
 Db 1 raarlgvrrxxrfgrvgr 18

## RESULT 13

AAW18147  
 ID AAW18147 standard; peptide; 18 AA.

XX AAW18147;

XX 11-AUG-1997 (first entry)

XX Cationic, antimicrobial, virus-neutralising protegrin PC-49.

XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;

XX Candida albicans; gram-negative bacteria; STD;

XX sexually transmitted disease; HIV-1; Chlamydia trachomatis;

XX Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;

XX food.

XX Synthetic.

XX Key Location/Qualifiers  
 FT Modified-site 1  
 FT /note= "Acylated"

FT Disulfide-bond 6..15

FT /note= "results in bullet form peptide"

FT Modified-site 18

FT /note= "Amidated"

FT W09637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

XX 07-JUL-1995; 95US-0499523.

XX 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX Harwig SSL, Kokryakov VN, Lehrer RI;

XX WPI; 1997-033984/03.

XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
 PT useful for the treatment of microbial infection, as food  
 PT preservatives and in eye care solutions

XX PS Claim 6; Page 64; 106pp; English.

XX CC The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. Peptides of this formula are designated protegrins and  
 CC are useful as anti-bacterial, anti-viral and anti-fungal agents in  
 CC plants and animals. The protegrins confer resistance to microbial or  
 CC viral infection in plants by preventing the growth of a virus or microbe  
 CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
 CC are particularly useful for the treatment of sexually transmitted  
 CC disease caused by microorganisms e.g. *Candida albicans*, HIV-1, *Chlamydia*  
 CC *trachomatis*, *Treponema pallidum* and *Neisseria gonorrhoeae*. They can also  
 CC be used in eye care solutions and as preservatives for food. The  
 CC protegrins are more effective under physiological conditions (e.g. in  
 CC the presence of serum) than certain antibiotics and are non-toxic to the  
 CC cells of higher organisms.

XX CC Sequence 18 AA;

Query Match 59.6%; Score 53; DB 18; Length 18;  
 Best Local Similarity 61.1%; Pred. No. 0.031;  
 Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18  
 ||||| : ||| : |||  
 Db 1 rggrlcwarrrrvcvgr 18

RESULT 14  
 AAW18148

ID AAW18148 standard; peptide; 18 AA.

XX AC AAW18148;

XX DT 11-AUG-1997 (first entry)

XX DE Cationic, antimicrobial, virus-neutralising protegrin PC-50.

XX KW Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW *Candida albicans*; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; *Chlamydia trachomatis*;  
 KW *Treponema pallidum*; *Neisseria gonorrhoeae*; eye care; preservative;  
 KW food.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1

FT Disulfide-bond /note= "Acylated"

FT Disulfide-bond 6..15

FT Modified-site /note= "results in bullet form peptide"

FT Modified-site 18

FT /note= "Amidated"

PN W09637508-A1.

XX PD 28-NOV-1996.

XX PF 24-MAY-1996; 96WO-US07594.

XX PR 07-JUL-1995; 95US-0499523.

PR XX 26-MAY-1995; 95US-0451832.

PA (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX PI Harwig SSL, Kokryakov VN, Lehrer RI;

XX DR WPI; 1997-033984/03.

XX PT Cationic, antimicrobial, virus-neutralising protegrin peptide(s)  
 PT useful for the treatment of microbial infection, as food  
 PT preservatives and in eye care solutions

XX PS Claim 6; Page 64; 106pp; English.

XX CC The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. Peptides of this formula are designated protegrins and  
 CC are useful as anti-bacterial, anti-viral and anti-fungal agents in  
 CC plants and animals. The protegrins confer resistance to microbial or  
 CC viral infection in plants by preventing the growth of a virus or microbe  
 CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
 CC are particularly useful for the treatment of sexually transmitted  
 CC disease caused by microorganisms e.g. *Candida albicans*, HIV-1, *Chlamydia*  
 CC *trachomatis*, *Treponema pallidum* and *Neisseria gonorrhoeae*. They can also  
 CC be used in eye care solutions and as preservatives for food. The  
 CC protegrins are more effective under physiological conditions (e.g. in  
 CC the presence of serum) than certain antibiotics and are non-toxic to the  
 CC cells of higher organisms.

XX CC Sequence 18 AA;

Query Match 59.6%; Score 53; DB 18; Length 18;  
 Best Local Similarity 61.1%; Pred. No. 0.031;  
 Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGGRLAYRLLRFAIRVGR 18  
 ||||| : ||| : |||  
 Db 1 rggrlcwarrrrvcvgr 18

RESULT 15  
 AAW18149

ID AAW18149 standard; peptide; 18 AA.

XX AC AAW18149;

XX DT 11-AUG-1997 (first entry)

XX DE Cationic, antimicrobial, virus-neutralising protegrin PC-52.

XX KW Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW *Candida albicans*; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; *Chlamydia trachomatis*;  
 KW *Treponema pallidum*; *Neisseria gonorrhoeae*; eye care; preservative;  
 KW food.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1

FT Disulfide-bond /note= "Acylated"

FT Disulfide-bond 8..13



FT Modified-site /note= "results in kite form peptide"  
18 /note= "Amidated"

XX WO9637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

XX 07-JUL-1995; 95US-0499523.

XX 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX Harwig SSL, Kokryakov VN, Lehrer RI;

XX WPI; 1997-033984/03.

XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
PT useful for the treatment of microbial infection, as food  
PT preservatives and in eye care solutions

XX Claim 6; Page 64; 106pp; English.

XX The present sequence is a specifically claimed example of a peptide,  
CC recombinantly produced, corresponding to the generic formula:  
CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
CC acid or proline; A17 may be absent or a basic, neutral/polar,  
CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
CC esterified forms, all of which may contain a disulphide bond to give a  
CC cysteine bridge. Peptides of this formula are designated protegrins and  
CC are useful as anti-bacterial, anti-viral and anti-fungal agents in  
CC plants and animals. The protegrins confer resistance to microbial or  
CC viral infection in plants by preventing the growth of a virus or microbe  
CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
CC are particularly useful for the treatment of sexually transmitted  
CC disease caused by microorganisms e.g. Candida albicans, HIV-1, Chlamydia  
CC trachomatis, Treponema pallidum and Neisseria gonorrhoeae. They can also  
CC be used in eye care solutions and as preservatives for food. The  
CC protegrins are more effective under physiological conditions (e.g. in  
CC the presence of serum) than certain antibiotics and are non-toxic to the  
CC cells of higher organisms.

XX SQ Sequence 18 AA;

Query Match 59.6%; Score 53; DB 18; Length 18;  
Best Local Similarity 61.1%; Pred. No. 0.031;  
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 RGGRLAVKLLRFAIRVGR 18

Db 1 RGGRLAVKLLRFAIRVGR 18

Search completed: February 12, 2002, 12:30:30  
Job time: 363 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:35 ; Search time 232.64 Seconds  
(without alignments)  
11.317 Million cell updates/sec

Title: US-09-485-571-18

Perfect score: 89

Sequence: 1 EGGELSYSEEFVSUGE 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_prodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query %	Length	ID	Description
1	48	53.9	1425	5	Q9VIT9	Q9vit9 drosophila
2	48	53.9	1496	5	Q9NFV5	Q9ntv5 drosophila
3	47	52.8	1158	3	Q9UTR5	Q9utr5 schizosacch
4	46	51.7	518	9	Q9XJB0	Q9xb0 streptococc
5	46	51.7	521	9	Q9MCJ9	Q9mcj9 streptococc
6	45	50.6	137	13	Q9PSV7	Q9psv7 conger myri
7	44	49.4	117	11	Q9D596	Q9d596 mus musculu
8	44	49.4	183	4	O75231	O75231 homo sapien
9	44	49.4	446	2	Q9AJ15	Q9aj15 escherichia
10	44	49.4	548	5	Q9VJD1	Q9vjd1 drosophila
11	43	48.3	300	6	O46383	O46383 bison bison
12	43	48.3	322	1	Q9HND2	Q9hnd2 halobacteri
13	43	48.3	492	6	O46384	O46384 bos taurus
14	43	48.3	515	9	O64283	O64283 streptococc
15	43	48.3	515	9	O64294	O64294 streptococc
16	43	48.3	1199	6	O28139	O28139 bos taurus
17	42	47.2	145	10	O24553	O24553 zea diplope
18	42	47.2	156	10	O41718	O41718 zea diplope
19	42	47.2	210	2	Q9A1H7	Q9a1h7 streptococc

20	42	47.2	243	5	O01830	O01830 caenorhabdi
21	42	47.2	316	10	O49814	O49814 capsicum an
22	42	47.2	324	10	O9SV15	O9sv15 arabidopsis
23	42	47.2	455	5	Q9GRV1	Q9grv1 caenorhabdi
24	42	47.2	503	12	Q9WI35	Q9wi35 cauliflower
25	42	47.2	573	5	Q93569	Q93569 caenorhabdi
26	42	47.2	751	4	Q9PLU4	Q9plu4 homo sapien
27	42	47.2	5038	4	Q9NPK1	Q9npk1 homo sapien
28	41	46.1	226	2	O53827	O53827 mycobacteri
29	41	46.1	228	2	O9CBC6	O9cbc6 mycobacteri
30	41	46.1	242	2	O50034	O50034 mycobacteri
31	41	46.1	279	12	O11861	O11861 bartonella
32	41	46.1	290	2	P71234	P71234 escherichia
33	41	46.1	304	4	Q9BQX6	Q9bx6 homo sapien
34	41	46.1	309	10	Q9FV41	Q9fv41 tagetes ere
35	41	46.1	348	2	O67217	O67217 aquifex aeo
36	41	46.1	350	2	O9CNN8	O9cnn8 pasteurella
37	41	46.1	354	3	Q9HGQ6	Q9hqg6 saccharomyc
38	41	46.1	399	2	O9AOL7	O9aol7 streptococc
39	41	46.1	446	2	O52140	O52140 escherichia
40	41	46.1	446	2	O85632	O85632 escherichia
41	41	46.1	459	2	O9RSU7	O9rsu7 deinococcus
42	41	46.1	499	6	O9MYX8	O9myx8 sus scrofa
43	41	46.1	618	3	Q9Z200	Q9z200 emericella
44	41	46.1	800	2	O52998	O52998 escherichia
45	41	46.1	7962	4	Q10465	Q10465 homo sapien

#### ALIGNMENTS

RESULT 1

Q9VIT9 ID Q9VIT9 PRELIMINARY; PRT; 1425 AA.  
AC Q9VIT9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CG10363 PROTEIN.  
GN TEPIV OR CG10363.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Beeson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Rosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RT Science 287:2185-2195(2000).  
 RL EMBL; AE003663; AAF53826.1; -;  
 DR HSSP; P01024; 1C3D.  
 DR FlyBase; FBgn0041180; Tepiv.  
 DR InterPro; IPR002890; A2M\_N.  
 DR InterPro; IPR001599; Alpha\_2\_macrolobln.  
 DR InterPro; IPR002114; PTS\_Hpr\_ser.  
 DR Pfam; PF00207; A2M; 1.  
 DR Pfam; PF01835; A2M\_N; 3.  
 DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
 DR PROSITE; PS00589; PTS\_HPR\_SER; UNKNOWN\_1.  
 SQ SEQUENCE 1425 AA; 160514 MW; C4B92282D2605CBE CRC64;

Query Match 53.9%; Score 48; DB 5; Length 1425;  
 Best Local Similarity 62.5%; Pred. No. 46;  
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 EGGELSYSEEEFSVSV 16  
 DB 524 ETGEFRYTEETSVEV 539

RESULT 2  
 Q9NFV5 PRELIMINARY; PRT; 1496 AA.  
 ID Q9NFV5;  
 AC Q9NFV5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE TEPA PROTEIN PRECURSOR.  
 GN TEPIV OR TEPA OR CG10363.  
 OS *Drosophila melanogaster* (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Laqueux M., Perrodou E., Levashina E.A., Capovilla M., Hoffmann J.A.;  
 RA "Constitutive expression of a novel complement-like protein in Toll  
 RT and Jak gain-of-function mutants of *Drosophila*.";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ269541; CAB87810.1;  
 DR FlyBase; FBgn0041180; Tepiv.  
 DR InterPro; IPR002890; A2M\_N.  
 DR InterPro; IPR001599; Alpha\_2\_macrolobln.  
 DR InterPro; IPR002114; PTS\_Hpr\_ser.  
 DR Pfam; PF00207; A2M; 1.  
 DR Pfam; PF01835; A2M\_N; 3.  
 DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; UNKNOWN\_1.  
 DR PROSITE; PS00589; PTS\_HPR\_SER; UNKNOWN\_1.  
 KW Signal.  
 FT SIGNAL  
 SQ SEQUENCE 1496 AA; 168491 MW; C7FB0FEE5C90AA2F CRC64;

Query Match 53.9%; Score 48; DB 5; Length 1496;  
 Best Local Similarity 62.5%; Pred. No. 48;  
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 EGGELSYSEEEFSVSV 16  
 DB 595 ETGEFRYTEETSVEV 610

RESULT 3  
 Q9UTR5 PRELIMINARY; PRT; 1158 AA.  
 ID Q9UTR5;  
 AC Q9UTR5;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE PUTATIVE RHO1 GDP-GTP EXCHANGE PROTEIN.  
 GN SPAC1006.06.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA McDougall R.C., Rajandream M.A., Barrell B.G., Simmonds M.,  
 RA Churcher C.M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL132828; CAB60236.1;  
 DR InterPro; IPR000219; RhoGEF.  
 DR InterPro; IPR000591; DEP.  
 DR InterPro; IPR001180; CNH.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR003880; Phosphopant\_attach.  
 DR Pfam; PF00610; DEP; 1.  
 DR Pfam; PF00621; RhoGEF; 1.  
 DR Pfam; PF00780; CNH; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
 DR SMART; SM00036; CNH; 1.  
 DR SMART; SM00049; DEP; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00325; RhoGEF; 1.  
 SQ SEQUENCE 1158 AA; 130878 MW; 6FFE8244710D33B1 CRC64;

Query Match 52.8%; Score 47; DB 3; Length 1158;  
 Best Local Similarity 55.6%; Pred. No. 52;  
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEEFSVSV 18  
 DB 1122 EGGEILYSTETIPFSSGE 1139

RESULT 4  
 Q9XJB0 PRELIMINARY; PRT; 518 AA.  
 ID Q9XJB0;  
 AC Q9XJB0;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE PUTATIVE TAIL COMPONENT PROTEIN.  
 OS *Streptococcus thermophilus* bacteriophage DT1.  
 CC Viruses.  
 OX NCBI\_TaxID=90410;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DT1;  
 RX MEDLINE=99160757; PubMed=10049822;  
 RA Tremblay D.M., Moineau S.;  
 RT "Complete genomic sequence of the lytic bacteriophage DT1 of  
 RT *Streptococcus thermophilus*.";  
 RL Virology 255:63-76(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN-DT1;  
 RA Tremblay D.M., Moineau S.;  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF085222; AAD21893.1; -;  
 SQ SEQUENCE 518 AA; 58300 MW; 34D230523784CB3B CRC64;

Query Match 51.7%; Score 46; DB 9; Length 518;  
 Best Local Similarity 53.3%; Pred. No. 29;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 ELSYSEEEFVSVE 18  
 ||:| | ||: ||  
 Db 478 ELTYLSEPFSGIGE 492

RESULT 5  
 Q9MCJ9 PRELIMINARY; PRT; 521 AA.  
 AC Q9MCJ9;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE ORF34.  
 GN ORF34.  
 OS Streptococcus thermophilus bacteriophage 7201.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
 OX NCBI\_TaxID=112023;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20088830; PubMed=10620678;  
 RA Stanley E., Walsh L., van der Zwet A., Fitzgerald G.F.,  
 van Sinderen D.;  
 RT "Identification of four loci isolated from two Streptococcus  
 thermophilus phage genomes responsible for mediating bacteriophage  
 resistance";  
 RL FEMS Microbiol. Lett. 182:271-277(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Stanley E., Walsh L., Fitzgerald G.F., van Sinderen D.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF145054; AAF43526.1; -;  
 SQ SEQUENCE 521 AA; 58576 MW; AF513FA740013C7B CRC64;

Query Match 51.7%; Score 46; DB 9; Length 521;  
 Best Local Similarity 53.3%; Pred. No. 29;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 ELSYSEEEFVSVE 18  
 ||:| | ||: ||  
 Db 481 ELTYLSEPFSGIGE 495

RESULT 6  
 Q9PSV7 PRELIMINARY; PRT; 137 AA.  
 AC Q9PSV7;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE CONGER EEL GALECTIN (CONGERIN I).  
 GN PCON I.  
 OS Conger myriaster (Conger eel).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Congroidei;  
 OC Congridae; Conger.  
 OX NCBI\_TaxID=7943;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ozawa T., Ishii C.;  
 RT "Galectin from skin of Conger myriaster";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AE010276; BAA36385.1; -;  
 DR HSP; P26788; ICIL.  
 DR InterPro: IPR001079; Gal-bind lectin.  
 DR Pfam: PF00337; Gal-bind lectin; 1.  
 DR SMART: SM00276; GLECT; 1.  
 DR PROSITE: PS00309; GALACTIN; 1.  
 SQ SEQUENCE 137 AA; 15448 MW; 689EB34F7E39B4F7 CRC64;

Query Match 50.6%; Score 45; DB 13; Length 137;  
 Best Local Similarity 52.9%; Pred. No. 8.5;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVE 18  
 ||:| | ||: ||  
 Db 21 GGFNNSPQRFVNVGE 37

RESULT 7  
 Q9D596 PRELIMINARY; PRT; 117 AA.  
 AC Q9D596;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE 4930488L21RIK PROTEIN.  
 GN 4930488L21RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,  
 Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sakai H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK015647; BAB29915.1; -;  
 DR MGD; MGI:1923059; 4930488L21RIK.  
 SQ SEQUENCE 117 AA; 12868 MW; 665940B7EF891419 CRC64;

Query Match 49.4%; Score 44; DB 11; Length 117;  
 Best Local Similarity 53.3%; Pred. No. 10;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVE 16  
 | - ||:| | ||: |  
 Db 46 GGRLLSHQEFSTKL 60

RESULT 8  
 O75231 PRELIMINARY; PRT; 183 AA.  
 ID O75231  
 AC O75231;

DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE WUGSC:H00905J08.3 PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cordes M., Wollam C., Carter T.;  
 RT "The sequence of Homo sapiens PAC clone DJ0905J08."  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005189; AAC25527.1; -;  
 DR HSP; P00518; IPHK.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PROSITE; P85011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; transferase.  
 FT NON\_TER 1  
 SQ SEQUENCE 183 AA; 20911 MW; 17E7040069D3E842 CRC64;

Query Match 49.4%; Score 44; DB 4; Length 183;  
 Best Local Similarity 90.0%; Pred. No. 18;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LYSSEEEFSV 14  
 DB 49 LYSSEEEFDV 58

RESULT 9  
 Q9AJ15 PRELIMINARY; PRT; 446 AA.  
 AC Q9AJ15;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DE ESCN.  
 GN ESCN.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RDEC-1;  
 RX MEDLINE=98254123; PubMed=9593291;  
 RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,  
 RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;  
 RT "The complete sequence of the locus of enterocyte effacement (LEE)  
 RT from enteropathogenic Escherichia coli E2348/69."  
 RL Mol. Microbiol. 28:1-4 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RDEC-1;  
 RX MEDLINE=21153569; PubMed=11254564;  
 RA Zhu C., Agin T.S., Elliott S.J., Johnson L.A., Thate T.E., Kaper J.B.,  
 RA Boedeker E.C.;  
 RT "Complete Nucleotide Sequence and Analysis of the Locus of Enterocyte  
 RT Effacement from Rabbit Diarrheagenic Escherichia coli RDEC-1";  
 RL Infect. Immun. 69:2107-2115 (2001).  
 DR EMBL; AF200363; AAK26715.1; -;  
 SQ SEQUENCE 446 AA; 48804 MW; DD782F98DD00F6632 CRC64;

Query Match 49.4%; Score 44; DB 2; Length 446;  
 Best Local Similarity 52.9%; Pred. No. 52;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGELSYSEEEFSVSVC 18  
 DB 92 GQWLSYQGEFFKIRVGD 108

RESULT 10  
 Q9VJDI PRELIMINARY; PRT; 548 AA.

ID Q9VJDI;  
 AC Q9VJDI;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE CG6453 PROTEIN.  
 GN CG6453.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer G.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Riden-Kiamos I., Saunders R.D.C., Scheeler F., Smith H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Shen T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodman T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195 (2000).  
 CC -!- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A  
 CC (LDLRA) DOMAIN.  
 CC -!- SIMILARITY: TO EF-HAND FAMILY.  
 CC EMBL; AF003655; AAF53621.1; -;  
 DR HSSP; P01130; 1LDL.  
 DR FlyBase; FBgn0032643; CG6453.

DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000886; ER\_target.  
 DR InterPro: IPR002172; LDL\_recept\_A.  
 DR Pfam: PF00036; efhand; 1.  
 DR SMART: SM00192; LDLa; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
 DR PROSITE: PSS0068; LDLRA\_2; 1.  
 KW Calcium-binding; Glycoprotein.  
 SQ SEQUENCE 548 AA; 61539 MW; 4F486B724D64732E CRC64;

Query Match 49.4%; Score 44; DB 5; Length 548;  
 Best Local Similarity 55.6%; Pred. No. 66;  
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVSVGE 18  
 |||||:|||||  
 Db 350 EGEEDQYDDEPFGVGE 367

RESULT 11  
 O46383 PRELIMINARY; PRT; 300 AA.  
 ID O46383;  
 AC O46383;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE NA-CA+K EXCHANGER (FRAGMENT).  
 GN BISNCKX.

OS Bison bison (American bison).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bison.  
 OX NCBI\_TaxID=9901;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98138491; PubMed=9478004;  
 RA Tucker J.E., Winkfein R.J., Cooper C.B., Schnetkamp P.P.;  
 RT "cDNA cloning of the human retinal rod Na-Ca + K exchanger: comparison  
 with a revised bovine sequence";  
 RL Invest. Ophthalmol. Vis. Sci. 39:435-440(1998).  
 DR EMBL: AF025480; AAC13320.1; -;  
 FT NON\_TER 1  
 FT NON\_TER 300 300  
 SQ SEQUENCE 300 AA; 31671 MW; 2BE592DA5AB9781E CRC64;

Query Match 48.3%; Score 43; DB 6; Length 300;  
 Best Local Similarity 44.4%; Pred. No. 47;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVSVGE 18  
 |||||:|||||  
 Db 73 EGGEVKGEDEGEIQAGE 90

RESULT 12  
 Q9HND2 PRELIMINARY; PRT; 322 AA.  
 ID Q9HND2  
 AC Q9HND2;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE ELECTRON TRANSFER FLAVOPROTEIN SUBUNIT ALPHA.  
 GN ETTA OR VNG2151G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;

RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL: AE005103; AAC20288.1; -;  
 DR InterPro: IPR001308; ETF\_alpha.  
 DR Pfam: PF00766; ETF\_alpha; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 322 AA; 33466 MW; B86AFDB6A5A8E1E3 CRC64;

Query Match 48.3%; Score 43; DB 1; Length 322;  
 Best Local Similarity 50.0%; Pred. No. 51;  
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVG 17  
 |||||:|||||  
 Db 192 GGVDVADIADAEFLVSVG 207

RESULT 13  
 O46384 PRELIMINARY; PRT; 492 AA.  
 ID O46384  
 AC O46384;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE NA-CA+K EXCHANGER (FRAGMENT).  
 GN BOSNCKX.

OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tucker J.E., Winkfein R.J., Cooper C.B., Schnetkamp P.P.M.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF025664; AAB88884.1; -;  
 DR InterPro: IPR002613; Na\_Ca\_Ex.  
 DR Pfam: PF01699; Na\_Ca\_Ex; 1.  
 FT NON\_TER 1  
 FT NON\_TER 492 492  
 SQ SEQUENCE 492 AA; 52337 MW; 7941A055D5C34B29 CRC64;

Query Match 48.3%; Score 43; DB 6; Length 492;  
 Best Local Similarity 44.4%; Pred. No. 85;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVSVGE 18  
 |||||:|||||  
 Db 81 EGGEVKGEDEGEIQAGE 98

RESULT 14  
 O64283 PRELIMINARY; PRT; 515 AA.  
 ID O64283  
 AC O64283;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE HYPOTHETICAL 58.5 KDA PROTEIN.  
 OS Streptococcus thermophilus bacteriophage Sfi21.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
 OX NCBI\_TaxID=64186;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE-98160788; PubMed-9499809;  
RA Desiere F., Lucchini S., Brussow H.;  
RT "Evolution of Streptococcus thermophilus bacteriophage genomes by  
RT modular exchanges followed by point mutations and small deletions and  
RT insertions.";  
RL Virology 241:345-356(1998).  
RL EMBL: AF115103; AAC39282.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 515 AA; 58543 MW; 10DB11E438C157B7 CRC64;

Query Match 48.3%; Score 43; DB 9; Length 515;  
Best Local Similarity 53.3%; Pred. No. 90;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 4 ELSYSEEEFSVSUGE 18  
||:| | | | |  
Db 475 ELTYLSEPFSGTGE 489

RESULT 15  
O64294  
ID O64294 PRELIMINARY; PRT; 515 AA.  
AC O64294;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE HYPOTHETICAL 58.5 KDA PROTEIN.  
OS Streptococcus thermophilus bacteriophage Sfil9.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
OX NCBI\_TaxID=72638;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98160788; PubMed-9499809;  
RA Desiere F., Lucchini S., Brussow H.;  
RT "Evolution of Streptococcus thermophilus bacteriophage genomes by  
RT modular exchanges followed by point mutations and small deletions and  
RT insertions.";  
RL Virology 241:345-356(1998).  
RL EMBL: AF115102; AAC39296.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 515 AA; 58475 MW; 09DE9B6AC7A480D CRC64;

Query Match 48.3%; Score 43; DB 9; Length 515;  
Best Local Similarity 53.3%; Pred. No. 90;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 4 ELSYSEEEFSVSUGE 18  
||:| | | | |  
Db 475 ELTYLSEPFSGTGE 489

Search completed: February 12, 2002, 12:38:36  
Job time: 749 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:48 ; Search time 67.2 seconds  
(without alignments)  
9.821 Million cell updates/sec

Title: US-09-485-571-18

Perfect score: 89

Sequence: 1 EGGELSYSEEEFSVSGE 18

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	50.6	135	1	LEGL_CONMY
2	45	50.6	233	1	EBSD_ENTFA
3	45	50.6	380	1	TRMU_DEIRA
4	44	49.4	414	1	S17A_HUMAN
5	43	48.3	168	1	PLAS_POPNI
6	43	48.3	191	1	YP6_AGRU
7	43	48.3	956	1	SYL_AQUAE
8	42	47.2	820	1	SYFE_DEIRA
9	42	47.2	5032	1	RYNR_HUMAN
10	41	46.1	155	1	PLAS_HORVU
11	41	46.1	168	1	PLAT_POPNI
12	41	46.1	318	1	KPR2_YEAST
13	41	46.1	355	1	KPR4_YEAST
14	41	46.1	360	1	LPXE_HELPY
15	41	46.1	448	1	CATE_BACFI
16	41	46.1	529	1	YB89_YEAST
17	41	46.1	812	1	FAED_ECOLI
18	40	44.9	99	1	PLAS_CUCPE
19	40	44.9	273	1	SIXI_MOUSE
20	40	44.9	284	1	SIXI_MOUSE
21	40	44.9	397	1	S17A_RABIT
22	40	44.9	442	1	YQHE_BACSU
23	40	44.9	497	1	TRXB_HUMAN
24	40	44.9	499	1	TRXB_BOVIN
25	40	44.9	852	1	CSG_HALHA
26	39.5	44.4	825	1	XFP_SCHPO
27	39	43.8	99	1	PLAS_MERPE
28	39	43.8	99	1	PLAS_SOLTU
29	39	43.8	149	1	PG1_PIG
30	39	43.8	170	1	PLAS_LYCES
31	39	43.8	241	1	CTR3_YEAST
32	39	43.8	286	1	TOAL_YEAST
33	39	43.8	330	1	YJBN_ECOLI

#### ALIGNMENTS

RESULT 1  
LEGL\_CONMY STANDARD; PRT; 135 AA.

AC P26788;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE CONGERIN I (BETA-GALACTOSIDE-BINDING LECTIN I).

OS Conger myriaster (Conger eel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Congroidei;

OC Congridae; Conger.

OX NCBI\_TaxID=7943;

RN [1]

RP SEQUENCE.

RC TISSUE=Skin mucus;

RX MEDLINE=92256465; PubMed=1581341;

RA Muramoto K., Kamiya H.;

RT "The amino-acid sequence of a lectin from conger eel, Conger

myriaster, skin mucus.";

RL Biochim. Biophys. Acta 1116:129-136(1992).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).

RX MEDLINE=20015373; PubMed=10545323;

RA Shirai T., Mitsuyama C., Niwa Y., Matsui Y., Hotta H., Yamane T.,

Kamiya H., Ishii C., Ogawa T., Muramoto K.;

RT "High-resolution structure of the conger eel galectin, congerin I, in

lactose-liganded and ligand-free forms: emergence of a new structure

class by accelerated evolution.";

RL Structure 7:1223-1233(1999).

CC -|- FUNCTION: THIS PROTEIN BINDS BETA-GALACTOSIDE. ITS PHYSIOLOGICAL

CC -|- FUNCTION IS NOT YET KNOWN.

CC -|- SUBUNIT: HOMODIMER.

CC -|- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.

DR PIR; S21102; S21102

DR PDB; 1C1F; 08-OCT-99.

DR PDB; 1C1F; 08-OCT-99.

DR InterPro; IPR001079; Gal-bind\_lectin.

DR Pfam; PF00337; Gal-bind\_lectin; 1.

DR SMART; SM00276; GLECT; 1.

DR PROSITE; PS00309; GALAPTIN; 1.

KW Galaptin; Lectin; Acetylation; 3D-structure.

FT MOD\_RES 1 ACETYLATION.

FT BINDING 70 76 BETA-GALACTOSIDE (POTENTIAL).

SQ SEQUENCE 135 AA; 15204 MW; 3AEC7E7E39BCE3B CRC64;

Query Match 50.6%; Score 45; DB 1; Length 135;

Best Local Similarity 52.9%; Pred. No. 1.9;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEEFSVSGE 18

DB 20 GGFINNSPQRFVNVGE 36

```

RESULT 2
EBSD_ENTFA
ID EBSD_ENTFA STANDARD; PRT; 253 AA.
AC P36923;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE 3-DEHYDROQUINATE DEHYDRATASE (EC 4.2.1.10) (3-DEHYDROQUINASE)
DE (TYPE I DHQASE).
GN EBSD.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OG1SSP;
RA MEDLINE=94042918; PubMed=8226689;
RX Bensing B.A., Dunny G.M.;
RT "Cloning and molecular analysis of genes affecting expression of
RT binding substance, the recipient-encoded receptor(s) mediating mating
RT aggregate formation in Enterococcus faecalis."
RL J. Bacteriol. 175:7421-7429(1993).
CC -!- CATALYTIC ACTIVITY: 3-DEHYDROQUINATE = 3-DEHYDROSHIKIMATE + H(2)O.
CC -!- PATHWAY: THIRD STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -!- SIMILARITY: BELONGS TO THE TYPE-I 3-DEHYDROQUINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L23802; AAC36854.1; -
CC InterPro; IPR001381; DHQinase_I.
CC Pfam; PF01487; DHQinase_I.
CC PROSITE; PS01028; DEHYDROQUINASE_I; 1.
CC Aromatic amino acid biosynthesis; Lyase.
FT ACT_SITE 143 143 BY SIMILARITY.
FT ACT_SITE 170 170 FORMS A SCHIFF-BASE INTERMEDIATE
FT (BY SIMILARITY).
SQ SEQUENCE 253 AA; 28085 MW; F8F1436A80906B02 CRC64;

Query Match 50.6%; Score 45; DB 1; Length 253;
Best Local Similarity 58.3%; Pred. No. 3.6;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGGELSYSEEF 12
    |||||:|:|:|
DB 86 EGGEMAFSEENY 97

RESULT 3
TRMU_DEIRA
ID TRMU_DEIRA STANDARD; PRT; 380 AA.
AC Q9RTK1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE TRNA (5-METHYLAMINOMETHYL-2-THIOURIDYLATE)-METHYLTRANSFERASE
DE (EC 2.1.1.61).
GN TRMU OR DR1759.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RA MEDLINE=20036896; PubMed=10567266;

```

```

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RT Science 286:1571-1577(1999).
RL -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA = S-ADENOSYL-L-
CC HOMOCYSTEINE + TRNA CONTAINING 5-METHYLAMINOMETHYL-2-
CC THIOURIDYLATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRMU FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE002017; AAF11314.1; ALT_INIT.
CC TIGR; DR1759; -
CC Transferase; Methyltransferase; TRNA processing; Complete proteome.
KW SEQUENCE 380 AA; 42184 MW; 000160AFC980A53 CRC64;

Query Match 50.6%; Score 45; DB 1; Length 380;
Best Local Similarity 44.4%; Pred. No. 5.6;
Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVSUGE 18
    |||||:|:|:|
DB 336 EGFEFAEPQFAVAPGO 353

RESULT 4
S17A_HUMAN
ID S17A_HUMAN STANDARD; PRT; 414 AA.
AC Q9UE55;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SERINE/THREONINE KINASE 17A (EC 2.7.1.-) (DAP KINASE-RELATED
DE APOPTOSIS-INDUCING PROTEIN KINASE 1).
GN SPK17A OR DR4K1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF LYS-90.
RC TISSUE=Placenta, and Liver;
RX MEDLINE=99003259; PubMed=9786912;
RA Sanjo H., Kawai T., Akira S.;
RT "DRAKs, novel serine/threonine kinases related to death-associated
RT protein kinase that trigger apoptosis."
RL J. Biol. Chem. 273:29066-29071(1998)
CC -!- FUNCTION: PHOSPHORYLATES THE MYOSIN LIGHT CHAIN (MLC) AS AN
CC EXOGENOUS SUBSTRATE AND IS INVOLVED IN APOPTOTIC SIGNALING. THE C-
CC TERMINAL REGION PLAYS AN IMPORTANT ROLE IN KINASE ACTIVITY AND FOR
CC INITIATION OF APOPTOSIS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PLACENTA AND IN LOWER
CC LEVELS IN HEART, LUNG, SKELETAL MUSCLE, KIDNEY AND PANCREAS.
CC -!- PTM: AUTOPHOSPHORYLATED.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AB011420; BAA34126.1; -  
 DR HSSP; P00518; IPHK.  
 DR MIM; 604726; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam; PF00069; pkinase.1.  
 DR SMART; SM00220; S\_TRC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_ST; 1.  
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 KW Phosphorylation; Nuclear protein; Apoptosis.  
 FT DOMAIN 61 321 PROTEIN\_KINASE.  
 FT NP\_BIND 67 75 ATP (BY SIMILARITY).  
 FT BINDING 90 90 ATP (BY SIMILARITY).  
 FT MUTAGEN 90 90 K->A: LOSS OF ACTIVITY.  
 SO SEQUENCE 414 AA; 46559 MW; 0C140290438C2A1A CRC64;

Query Match 49.4%; Score 44; DB 1; Length 414;  
 Best Local Similarity 90.0%; Pred. No. 9;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 LSYSEEFVS 14  
 I I I I I I I I I  
 Db 280 LSYSEEFVDV 289

## RESULT -5

PLAS\_POPNI  
 ID PLAS\_POPNI STANDARD; PRT; 168 AA.  
 AC P00299;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PLASTOCYANIN A, CHLOROPLAST PRECURSOR.  
 GN PETE.  
 OS Populus nigra (Lombardy poplar).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Malpighiales; Salicaceae; Populus.  
 OX NCBI\_TaxID=3691;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, ITALICA; TISSUE=Leaf;  
 RA Reichert J., Jenzelewski V., Hachnel W.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 RP [2]  
 RP SEQUENCE OF 70-168.  
 RC STRAIN=CV, ITALICA;  
 RA Ambler R.P.;  
 RL Unpublished results, cited by:  
 RL Freeman H.C.;  
 RL J. Proc. Royal Soc. N.S. Wales 112:45-62(1979).  
 RP [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=84135769; PubMed=6698995;  
 RA Garrett T.P.J., Clingeffer D.J., Guss J.M., Rogers S.J.,  
 RA Freeman H.C.;  
 RT "The crystal structure of poplar plastocyanin at 1.8-A resolution.  
 RT The geometry of the copper-binding site is created by the  
 RT polypeptide.";  
 RL J. Biol. Chem. 259:2822-2825(1984).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=84010876; PubMed=6620385;

RA Guss J.M., Freeman H.C.;  
 RT "Structure of oxidized poplar plastocyanin at 1.6-A resolution.";  
 RL J. Mol. Biol. 169:521-563(1983).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
 RA Colman P.M., Freeman H.C., Guss J.M., Murata M., Norris V.A.,  
 RA Ramshaw J.A.M., Venkatappa M.P.;  
 RT "X-ray crystal structure analysis of plastocyanin at 2.7-A  
 RT resolution.";  
 RL Nature 272:319-324(1978).  
 CC -1- FUNCTION: PLASTOCYANIN PARTICIPATES IN ELECTRON TRANSFER BETWEEN  
 CC P700 AND THE CYTOCHROME B/F COMPLEX IN PHOTOSYSTEM I.  
 CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID  
 CC MEMBRANE SURFACE IN CHLOROPLASTS.  
 CC -1- SIMILARITY: THERE ARE 12 DIFFERENCES BETWEEN THE SEQUENCE OF  
 CC POPLAR PLASTOCYANINS A AND B.  
 CC -1- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z50185; CAA90564.1; -  
 DR PIR; A00309; CUPX.  
 DR PDB; 2PCY; 15-JAN-86.  
 DR PDB; 3PCY; 25-APR-86.  
 DR PDB; 4PCY; 15-JAN-91.  
 DR PDB; 5PCY; 15-JAN-91.  
 DR PDB; 6PCY; 15-JAN-91.  
 DR PDB; 1PLC; 31-OCT-93.  
 DR PDB; 1PNC; 31-JAN-94.  
 DR PDB; 1PND; 31-JAN-94.  
 DR Mendel; 12195; POPNI; Pete; 1.  
 DR InterPro; IPR001235; Copper\_blue.  
 DR InterPro; IPR000923; Copper\_blue.  
 DR Pfam; PF00127; copper-bind; 1.  
 DR PRINTS; PR00156; COPPERBLUE.  
 DR PRINTS; PR00157; PLASTOCYANIN.  
 DR PRODOM; PD001235; Copper\_blue; 1.  
 DR PROSITE; PS00196; COPPER\_BLUE; 1.  
 DR Chloplast; Electron transport; Copper; Thylakoid; Membrane;  
 KW Transitt peptide; Multigene family; 3D-structure.  
 FT TRANSIT 1 69 CHLOROPLAST.  
 FT CHAIN 70 168 PLASTOCYANIN A.  
 FT DOMAIN 70 168 PLASTOCYANIN-LIKE.  
 FT METAL 106 106 COPPER.  
 FT METAL 153 153 COPPER.  
 FT METAL 156 156 COPPER.  
 FT METAL 161 161 COPPER.  
 FT STRAND 71 74  
 FT TURN 77 78  
 FT STRAND 83 84  
 FT STRAND 87 90  
 FT TURN 92 93  
 FT STRAND 95 100  
 FT STRAND 106 106  
 FT STRAND 109 110  
 FT TURN 112 113  
 FT TURN 117 118  
 FT TURN 121 123  
 FT HELIX 124 124  
 FT TURN 132 132  
 FT STRAND 135 136  
 FT TURN 138 142  
 FT STRAND 147 152  
 FT STRAND 154 156  
 FT HELIX 157 160  
 FT STRAND 162 167  
 SQ SEQUENCE 168 AA; 17020 MW; 901B21A7573DBF82 CRC64;



CC PYROPHOSPHATE + L-PHENYALANYL-TRNA(PHE).  
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA  
 CC CHAIN FAMILY, SUBFAMILY 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE002066; AAF11903.1; .  
 CC TIGR: DR2357; .  
 CC DR TIGR: DR2357; .  
 CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 CC Complete proteome.  
 CC SQ SEQUENCE 820 AA; 86984 MW; 1BE394EB78F7493E CRC64;  
 CC -----  
 CC Query Match 47.2%; Score 42; DB 1; Length 820;  
 CC Best Local Similarity 58.8%; Pred. No. 40;  
 CC Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 CC -----  
 CC QY 2 GGELSYSEEFVSVSGE 18  
 CC ||||| | | | | |  
 CC Db 757 GGELLESVEPFDVTGE 773  
 CC -----  
 CC RESULT 9  
 CC ID RYNR\_HUMAN STANDARD; PRT; 5032 AA.  
 CC AC P21817;  
 CC DT 01-MAY-1991 (Rel. 18, Created)  
 CC DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 CC DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 CC DE RYANODINE RECEPTOR, SKELETAL MUSCLE (SKELETAL MUSCLE CALCIUM RELEASE  
 CC CHANNEL).  
 CC GN RYR1 OR RYDR.  
 CC OS Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC OX NCBI\_TaxID=9606;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 CC RC TISSUE-Skeletal muscle;  
 CC RX MEDLINE-92128959; PubMed-1774074;  
 CC RA Zorzato F., Fujii J., Otsu K., Phillips M., Green N.M., Lai F.A.,  
 CC RA Meissner G., MacLennan D.H.;  
 CC RT "Molecular cloning of cDNA encoding human and rabbit forms of the  
 CC RT Ca2+ release channel (ryanodine receptor) of skeletal muscle  
 CC RT sarcoplasmic reticulum.";  
 CC RL J. Biol. Chem. 265:2244-2256(1990).  
 CC RN [2]  
 CC RP VARIANT MH CYS-614.  
 CC RX MEDLINE-92128959; PubMed-1774074;  
 CC RA Gillard E.F., Otsu K., Fujii J., Khanna V.K., de Leon S.,  
 CC RA Dardamezi J., Britt B.A., Duff C.L., Worton R.G., McLennan D.H.;  
 CC RT "A substitution of cysteine for arginine 614 in the ryanodine  
 CC RT receptor is potentially causative of human malignant hyperthermia.";  
 CC RL Genomics 11:751-755(1991).  
 CC RN [3]  
 CC RP VARIANT MH ARG-248, AND VARIANTS CYS-471; LEU-1786 AND CYS-2059.  
 CC RX MEDLINE-92372020; PubMed-1354642;  
 CC RA Gillard E.F., Otsu K., Fujii J., Duff C., de Leon S., Khanna V.K.,  
 CC RA Britt B.A., Worton R.G., McLennan D.H.;  
 CC RT "Polymorphisms and deduced amino acid substitutions in the coding  
 CC RT sequence of the ryanodine receptor (RYR1) gene in individuals with  
 CC RT malignant hyperthermia.";  
 CC RL Genomics 13:1247-1254(1992).  
 CC RN [4]

RP VARIANT CCD HIS-2434.  
 RX MEDLINE-94035117; PubMed-8220422;  
 RA Zhang Y., Chen H.S., Khanna V.K., de Leon S., Phillips M.S.,  
 RA Schappert K., Britt B.A., Brownell A.K.W., MacLennan D.H.;  
 RT "A mutation in the human ryanodine receptor gene associated with  
 RT central core disease.";  
 RL Nat. Genet. 5:46-50(1993).  
 RN [5]  
 RN VARIANTS CCD CYS-163 AND MET-403.  
 RX MEDLINE-94035118; PubMed-8220423;  
 RA Quane K.A., Healy J.M.S., Keating K.E., Manning B.M., Couch F.J.,  
 RA Palmucci L.M., Doriguzzi C., Fagerlund T.H., Berg K., Ordling H.,  
 RA Bendixen D., Mortier W., Linz U., Muller C.R., McCarthy T.V.;  
 RA "Mutations in the ryanodine receptor gene in central core disease and  
 RT malignant hyperthermia.";  
 RL Nat. Genet. 5:51-55(1993).  
 RN [6]  
 RN VARIANT MH SER-522.  
 RP MEDLINE-95130087; PubMed-7829078;  
 RX Quane K.A., Keating K.E., Healy J.M.S., Manning B.M.,  
 RA Krivosic-Horber R., Krivosic I., Monnier N., Lunardi J.,  
 RA McCarthy T.V.;  
 RT "Mutation screening of the RYR1 gene in malignant hyperthermia:  
 RT detection of a novel Tyr to Ser mutation in a pedigree with  
 RT associated central cores.";  
 RL Genomics 23:236-239(1994).  
 RN [7]  
 RN VARIANT MH ARG-341.  
 RP MEDLINE-94282042; PubMed-8012359;  
 RX Quane K.A., Keating K.E., Manning B.M., Healy J.M.S., Monniers K.,  
 RA Heffron J.J.A., Lehane M., Heytons L., Krivosic-Horber R., Adnet P.,  
 RA Ellis F.R., Monnier N., Lunardi J., McCarthy T.V.;  
 RA "Detection of a novel common mutation in the ryanodine receptor gene  
 RT in malignant hyperthermia: Implications for diagnosis and  
 RT heterogeneity studies.";  
 RL Hum. Mol. Genet. 3:471-476(1994).  
 RN [8]  
 RN VARIANT MH ARG-2433.  
 RX MEDLINE-95152512; PubMed-7849712;  
 RA Keating K.E., Quane K.A., Manning B.M., Lehane M., Hartung E.,  
 RA Censier K., Urwyler A., Klausnitzer M., Muller C.R., Heffron J.J.A.,  
 RA McCarthy T.V.;  
 RT "Detection of a novel RYR1 mutation in four malignant hyperthermia  
 RT pedigrees.";  
 RL Hum. Mol. Genet. 3:1855-1858(1994).  
 RN [9]  
 RN VARIANT MH ARG-2433.  
 RX MEDLINE-95187158; PubMed-7881417;  
 RA Phillips M.S., Khanna V.K., de Leon S., Frodis W., Britt B.A.,  
 RA MacLennan D.H.;  
 RT "The substitution of Arg for Gly2433 in the human skeletal muscle  
 RT ryanodine receptor is associated with malignant hyperthermia.";  
 RL Hum. Mol. Genet. 3:2181-2186(1994).  
 RN [10]  
 RN VARIANT MH ARG-35.  
 RP MEDLINE-97219028; PubMed-9066328;  
 RX Lynch F.J., Krivosic-Horber R., Reyford H., Monnier N., Quane K.,  
 RA Adnet F., Haudecoeur G., Krivosic I., McCarthy T., Lunardi J.;  
 RT "Identification of heterozygous and homozygous individuals with the  
 RT novel FHR1 mutation Cys35Arg in a large kindred.";  
 RL Anesthesiology 86:620-626(1997).  
 RN [11]  
 RN VARIANT MH LEU-614.  
 RP MEDLINE-98051290; PubMed-9389851;  
 RX Quane K.A., Ordling H., Keating K.E., Manning B.M., Heine R.,  
 RA Bendixen D., Berg K., Krivosic-Horber R., Lettmann-Horn F.,  
 RA Fagerlund T., McCarthy T.V.;  
 RT "Detection of a novel mutation at amino acid position 614 in the  
 RT ryanodine receptor in malignant hyperthermia.";  
 RL Br. J. Anaesth. 79:332-337(1997).  
 RN [12]  
 RN VARIANT MH TRP-552.  
 RP MEDLINE-97284075; PubMed-9138151;  
 RX

CC -!- MISCELLANEOUS: THE CALCIUM RELEASE CHANNEL ACTIVITY RESIDES IN THE  
CC C-TERMINAL REGION WHILE THE REMAINING PART OF THE PROTEIN  
CC CONSTITUTES THE 'FOOT' STRUCTURE SPANNING THE JUNCTIONAL GAP  
CC BETWEEN THE SR AND THE T-TUBULE. IT IS POSSIBLE THAT THE FOOT  
CC STRUCTURE INTERACTS WITH THE CYTOPLASMIC REGION OF THE  
CC DIHYDROPYRIDINE RECEPTOR.  
CC -!- MISCELLANEOUS: RYANODINE IS AN ALKALOID THAT BINDS TO THE  
CC CA-RELEASE CHANNEL IN JUNCTIONAL SR AND MODULATES ITS ACTIVITY.  
CC -!- SIMILARITY: LOCAL & LOW WITH THE NICOTINIC ACETYLCHOLINE RECEPTOR  
CC (N-ACHR) SUBUNITS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use.  
CC -----  
Query Match 47.2%; Score 42; DB 1; Length 5032;  
Best Local Similarity 55.6%; Pred. No. 2.8e+02;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
QY 1 EGGELSYSEEEFVSVSVE 18  
      ||||| :||| || |  
Db 3731 EGGENGAEVEEVSFEE 3748  
-----  
RESULT 10  
PLAS\_HORVU STANDARD; PRT; 155 AA.  
ID AC AC08248;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PLASTOCYANIN PRECURSOR.  
GN PETE.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N. A.  
RC STRAIN-CV. BOMI.  
RA Nielsen O.S.; Gausing K.;  
RT "The precursor of barley plastocyanin: sequence of cDNA clones and  
RT gene expression in different tissues.";  
RT FEBS Lett. 225:159-162(1987).  
RN [2]  
RP SEQUENCE FROM N. A.  
RC STRAIN-CV. NK 1558;  
RX MEDLINE=94039081; PubMed=8223592;  
RA Nielsen P.; Gausing K.;  
RT "In vitro binding of nuclear proteins to the barley plastocyanin gene  
RT promoter region.";  
RL Eur. J. Biochem. 217:97-104(1993).  
CC -!- FUNCTION: PLASTOCYANIN PARTICIPATES IN ELECTRON TRANSFER BETWEEN  
CC P700 AND THE CYTOCHROME B/F COMPLEX IN PHOTOSYSTEM I.  
CC -!- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID  
CC MEMBRANE SURFACE IN CHLOROPLASTS.  
CC -!- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Y00704; CAA68696.1; -  
CC EMBL: Z28347; CAA82201.1; -  
CC FIR: S00206; S00206.  
CC HSSP: P00289; 2PCF.  
CC Mendel; 8616; HORVU; PetE;1.  
CC InterPro: IPR001235; Copper\_blue.

```

DR InterPro: IPR000923; Copper_blue1.
DR Pfam: PF00127; copper-bind; 1.
DR PRINTS: PR00156; COPPERBLUE.
DR PRINTS: PR00157; PLASTOCYANIN.
DR ProDom: PD001235; Copper_blue; 1.
DR PROSITE: PS00196; COPPER_BLUE; 1.
KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
KW Transit peptide.
FT TRANSIT 1 58 CHLOROPLAST.
FT CHAIN 59 155 PLASTOCYANIN.
FT DOMAIN 59 155 PLASTOCYANIN-LIKE.
FT METAL 95 95 COPPER (BY SIMILARITY).
FT METAL 140 140 COPPER (BY SIMILARITY).
FT METAL 143 143 COPPER (BY SIMILARITY).
FT METAL 148 148 COPPER (BY SIMILARITY).
FT VARIANT 120 120 T -> N (IN CV. NK 1558).
SQ SEQUENCE 155 AA; 15709 MW; DAA7EABE5F6F4F91 CRC64;

Query Match 46.1%; Score 41; DB 1; Length 155;
Best Local Similarity 47.1%; Pred. No. 9.8;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSQVGE 18
   : | | : : | | | |
Db 67 GGVLFEPNDFSVKAGE 83

RESULT 11
PLAT_POPNI
ID PLAT_POPNI STANDARD; PRT; 168 AA.
AC P11970;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PLASTOCYANIN B, CHLOROPLAST PRECURSOR.
GN PETE.
OS Populus nigra (Lombardy poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3691;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ITALICA; TISSUE=Leaf;
RA Reichert J., Jenzelewski V., Haehnel W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 70-168.
RC STRAIN=CV. ITALICA;
RA Dimitrov M.I., Egorov C.A., Donchev A.A., Atanasov B.P.;
RT "Complete amino acid sequence of poplar plastocyanin b.";
RL FEBS Lett. 226:17-22(1987).
CC -1- FUNCTION: PLASTOCYANIN PARTICIPATES IN ELECTRON TRANSFER BETWEEN
CC P700 AND THE CYTOCHROME B/F COMPLEX IN PHOTOSYSTEM I.
CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
CC MEMBRANE SURFACE IN CHLOROPLASTS.
CC -1- SIMILARITY: THERE ARE 12 DIFFERENCES BETWEEN THE SEQUENCE OF
CC POPLAR PLASTOCYANINS A AND B.
CC -1- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. Usage by and for commercial
or send an email to license@sib-sib.ch).
-----
DR EMBL: Z50186; CAA90565.1; -.
DR PIR: S00210; S00210.
DR HSP: P00299; 4PCY.
DR Mendel: 12196; POPni; Pete; 2.

```

```

DR InterPro: IPR001235; Copper_blue.
DR InterPro: IPR000923; Copper_blue1.
DR Pfam: PF00127; copper-bind; 1.
DR PRINTS: PR00156; COPPERBLUE.
DR PRINTS: PR00157; PLASTOCYANIN.
DR ProDom: PD001235; Copper_blue; 1.
DR PROSITE: PS00196; COPPER_BLUE; 1.
KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
KW Transit peptide; Multigene family.
FT TRANSIT 1 69 CHLOROPLAST.
FT CHAIN 70 168 PLASTOCYANIN B.
FT DOMAIN 70 168 PLASTOCYANIN-LIKE.
FT METAL 106 106 COPPER.
FT METAL 153 153 COPPER.
FT METAL 156 156 COPPER.
FT METAL 161 161 COPPER.
SQ SEQUENCE 168 AA; 16981 MW; F20DA6EA2038AEEA CRC64;

Query Match 46.1%; Score 41; DB 1; Length 168;
Best Local Similarity 44.4%; Pred. No. 11;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVSQVGE 18
   : | | : : | | | |
Db 77 DDGSLAFVPEFSVPAGE 94

RESULT 12
KPR2_YEAST
ID KPR2_YEAST STANDARD; PRT; 318 AA.
AC P38620;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE RIBOSE-PHOSPHATE PYROPHOSPHOKINASE 2 (EC 2.7.6.1) (PHOSPHORIBOSYL
DE PYROPHOSPHATE SYNTHETASE 2).
GN PRPS2 OR PRS2 OR PRS OR YER099C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SKO2N;
RX MEDLINE=95084630; PubMed=7992503;
RA Carter A.T., Narbad A., Pearson B.M., Beck K.-F., Logghe M.,
RA Contreras R., Schweizer M.;
RT "Phosphoribosylpyrophosphate synthetase (PRS): a new gene family in
RL Yeast 10:1031-1044(1994).
[2]
RN ERRATUM.
RA Carter A.T., Narbad A., Pearson B.M., Beck K.-F., Logghe M.,
RA Contreras R., Schweizer M.;
RL Yeast 11:191-191(1995).
[3]
RN SEQUENCE FROM N.A.
RA Gerhardt H., Switzer R.L., Smith J.M., Hove-Jensen B.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Benito A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman E., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shooren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + D-RIBOSE 5-PHOSPHATE -> AMP +
CC 5-PHOSPHO-ALPHA-D-RIBOSE 1-DIPHOSPHATE.

```

CC CC -1- PATHWAY: THIS ENZYME IS UTILIZED BY BOTH THE DE NOVO & THE SALVAGE  
 CC CC PATHWAYS BY WHICH ENDOGENOUSLY FORMED OR EXOGENOUSLY ADDED  
 CC CC PYRIMIDINE, PURINE, OR PYRIDINE BASES ARE CONVERTED TO THE  
 CC CC CORRESPONDING RIBONUCLEOSIDE MONOPHOSPHATES.  
 CC CC -1- SIMILARITY: BELONGS TO THE RIBOSE-PHOSPHATE PYROPHOSPHOKINASE  
 CC CC FAMILY.  
 CC CC  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC CC EMBL; X75075; CAA52969.1; -;  
 CC CC EMBL; X74414; CAA52436.1; -;  
 CC CC EMBL; U18839; AAB64654.1; -;  
 CC CC SGD; S37225; S37225;  
 CC CC SGD; S0000901; PRS2.  
 CC CC InterPro: IPR000842; PRPP\_synthetase.  
 CC CC InterPro: IPR000836; Pribosyltran.  
 CC CC Pfam: PF00156; Pribosyltran; 1.  
 CC CC PROSITE: PS00114; PRPP\_SYNTHETASE; 1.  
 CC CC Nucleotide biosynthesis; Transferase; Kinase; Magnesium;  
 CC CC Multigene family.  
 CC CC FT METAL 132 132 MAGNESIUM (POTENTIAL).  
 CC CC FT METAL 134 134 MAGNESIUM (POTENTIAL).  
 CC CC FT METAL 143 143 MAGNESIUM (POTENTIAL).  
 CC CC FT METAL 147 147 MAGNESIUM (POTENTIAL).  
 CC CC SQ SEQUENCE 318 AA; 34765 MW; 8B970E98084F5D71 CRC64;  
 CC CC  
 CC CC Query Match 46.1%; Score 41; DB 1; Length 318;  
 CC CC Best Local Similarity 50.0%; Pred. No. 21;  
 CC CC Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 CC CC  
 CC CC QY 3 GELSYSEEEFVSUGE 18  
 CC CC | | | | | | | | | |  
 CC CC Db 33 GYQYSNKETSVTIGE 48  
 CC CC  
 CC CC RESULT 13  
 CC CC ID KPR4\_YEAST STANDARD; PRT; 355 AA.  
 CC CC AC P38063;  
 CC CC DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 CC CC DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 CC CC DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 CC CC DE PROBABLE RIBOSE-PHOSPHATE PYROPHOSPHOKINASE 4 (EC 2.7.6.1)  
 CC CC DE (PHOSPHORIBOSYL PYROPHOSPHATE SYNTHETASE 4).  
 CC CC GN PRPS4 OR PRS4 OR YBL068W OR YBL0619.  
 CC CC OS Saccharomyces cerevisiae (Baker's yeast).  
 CC CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC CC ON NCBI\_TaxID=4932;  
 CC CC RX [1]  
 CC CC RP SEQUENCE FROM N.A.  
 CC CC RC STRAIN=S288C;  
 CC CC RA Contreras R., Fiers W., Logghe M., Molemans F.;  
 CC CC RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC CC CC -1- CATALYTIC ACTIVITY: ATP + D-RIBOSE 5-PHOSPHATE = AMP +  
 CC CC 5-PHOSPHO-ALPHA-D-RIBOSE 1-DIPHOSPHATE.  
 CC CC -1- SIMILARITY: BELONGS TO THE RIBOSE-PHOSPHATE PYROPHOSPHOKINASE  
 CC CC FAMILY.  
 CC CC  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC

CC CC EMBL; Z35829; CAA84888.1; -;  
 CC CC PIR; S45804; S45804.  
 CC CC SGD; S0000164; PRS4.  
 CC CC InterPro: IPR000842; PRPP\_synthetase.  
 CC CC InterPro: IPR000836; Pribosyltran.  
 CC CC Pfam: PF00156; Pribosyltran; 1.  
 CC CC PROSITE: PS00114; PRPP\_SYNTHETASE; 1.  
 CC CC Nucleotide biosynthesis; Transferase; Kinase; Magnesium;  
 CC CC Multigene family.  
 CC CC FT METAL 168 168 MAGNESIUM (POTENTIAL).  
 CC CC FT METAL 170 170 MAGNESIUM (POTENTIAL).  
 CC CC FT METAL 179 179 MAGNESIUM (POTENTIAL).  
 CC CC FT METAL 183 183 MAGNESIUM (POTENTIAL).  
 CC CC SQ SEQUENCE 355 AA; 39059 MW; 41EBD027E539E46 CRC64;  
 CC CC  
 CC CC Query Match 46.1%; Score 41; DB 1; Length 355;  
 CC CC Best Local Similarity 50.0%; Pred. No. 24;  
 CC CC Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 CC CC  
 CC CC QY 3 GELSYSEEEFVSUGE 18  
 CC CC | | | | | | | | | |  
 CC CC Db 69 GYQYSNKETSVTIGE 84  
 CC CC  
 CC CC RESULT 14  
 CC CC LPXB\_HELPY STANDARD; PRT; 360 AA.  
 CC CC ID Q25537;  
 CC CC DT 15-DEC-1998 (Rel. 37, Created)  
 CC CC DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 CC CC DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 CC CC DE LIPID-A-DISACCHARIDE SYNTHASE (EC 2.4.1.182).  
 CC CC GN LPXB OR HP0867.  
 CC CC OS Helicobacter pylori (Campylobacter pylori).  
 CC CC OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 CC CC OC Helicobacter.  
 CC CC OX NCBI\_TaxID=210;  
 CC CC RN [1]  
 CC CC RP SEQUENCE FROM N.A.  
 CC CC RC STRAIN=26695 / ATCC 700392;  
 CC CC RX MEDLINE=97394467; PubMed=9252185;  
 CC CC RA Tomb J.-E., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 CC CC RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 CC CC RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 CC CC RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 CC CC RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 CC CC RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 CC CC RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
 CC CC RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 CC CC RA Venter J.C.;  
 CC CC RT "The complete genome sequence of the gastric pathogen Helicobacter  
 CC CC pylori".  
 CC CC RL Nature 388:539-547(1997).  
 CC CC CC -1- FUNCTION: CONDENSATION OF UDP-2,3-DIACYLGLUCOSAMINE AND  
 CC CC 2,3-DIACYLGLUCOSAMINE-1-PHOSPHATE TO FORM LIPID A DISACCHARIDE.  
 CC CC A PRECURSOR OF LIPID A, A PHOSPHORYLATED GLYCOLIPID THAT ANCHORS  
 CC CC THE LIPOPOLYSACCHARIDE TO THE OUTER MEMBRANE OF THE CELL  
 CC CC (BY SIMILARITY).  
 CC CC CC -1- CATALYTIC ACTIVITY: UDP-2,3-BIS(3-HYDROXYTETRADECANOYL)GLUCOSAMINE  
 CC CC + 2,3-BIS(3-HYDROXYTETRADECANOYL)-BETA-D-GLUCOSAMINYL 1-PHOSPHATE  
 CC CC -> UDP + 2,3-BIS(3-HYDROXYTETRADECANOYL)-D-GLUCOSAMINYL-1,6-BETA-  
 CC CC D-2,3-BIS(3-HYDROXYTETRADECANOYL)-BETA-D-GLUCOSAMINYL 1-PHOSPHATE.  
 CC CC -1- PATHWAY: LIPID A BIOSYNTHESIS.  
 CC CC -1- SIMILARITY: BELONGS TO THE LPXB FAMILY.  
 CC CC  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC



CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

CC EMBL; AE000596; AAD07909.1; -  
DR TIGR; HP0867; -  
DR InterPro; IPR003835; LpxB.  
DR Pfam; PF02684; LpxB; 1.  
KW Transferase; Glycosyltransferase; Lipid A biosynthesis;  
KW Lipid synthesis; Complete proteome.  
SQ SEQUENCE 360 AA; 41517 MW; 420421646FCE067D CRC64;

Query Match 46.1%; Score 41; DB 1; Length 360;

Best Local Similarity 57.1%; Pred. No. 24;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVS 14

|||:||||

DB 35 EGKEVLYSPREFSI 48

RESULT 15

CATE\_BACFI

ID CATE\_BACFI STANDARD; PRT; 448 AA.

AC P30266;

DT 01-APR-1993 (Rel. 25, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE CATALASE (EC 1.11.1.6) (FRAGMENT).

GN KATE OR KATA.

OS Bacillus firmus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=1399;

RN [1]

RP SEQUENCE FROM N.A.

RA Quirk P.G., Krulwich T.A.;

RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES

CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.

CC -!- CATALYTIC ACTIVITY: 2 H<sub>2</sub>O(2) = O(2) + 2 H(2)O.

CC -!- COFACTOR: HEME GROUP.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).

CC -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY. HP11 SUBFAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

CC EMBL; L02551; AAA22558.1; -.

DR PIR; S27490; S27490.

DR HSPG; P21179; IIPH.

DR InterPro; IPR002226; Catalase.

DR Pfam; PF00199; catalase; 1.

DR PRINTS; PR00067; CATALASE.

DR ProDom; PD000510; Catalase; 1.

DR PROSITE; PS00437; CATALASE\_1; 1.

DR PROSITE; PS00438; CATALASE\_2; 1.

KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.

FT ACT\_SITE 75 75 BY SIMILARITY.

FT ACT\_SITE 148 148 BY SIMILARITY.

FT BINDING 362 362 PROXIMAL HEME LIGAND (BY SIMILARITY).

FT NON\_TER 448 448

SQ SEQUENCE 448 AA; 51673 MW; 4D14048CB7023FD3 CRC64;

Query Match

Best Local Similarity 46.1%; Score 41; DB 1; Length 448;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Search completed: February 12, 2002, 12:39:49

Job time: 802 sec

Qy 5 LSYSEEFVS SVGE 18  
| | | | | |  
Db 33 LKVSDEFFSLKAGE 46



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:36 ; Search time 126.85 Seconds  
(without alignments)  
10.809 Million cell updates/sec

Title: US-09-485-571-18  
Perfect score: 89  
Sequence: 1 EGGELSYSEEEFSVSUGE 18  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	52.8	1158	T50454	probable rho1 GDP-
2	45	50.6	135	S21102	lectin - eel (Cong
3	45	50.6	253	D49539	hypothetical prote
4	45	50.6	413	H75357	tRNA (5-methylamin
5	43	48.3	168	1 CUPX	plastocyanin a pre
6	43	48.3	322	D84365	hypothetical prote
7	43	48.3	322	D0AG6T	electron transfer
8	43	48.3	515	T09272	probable tail comp
9	43	48.3	515	T09203	probable tail comp
10	43	48.3	956	G70327	isoleucine--tRNA 1
11	43	48.3	1199	S20969	Na+/Ca2+,K+-exchan
12	42	47.2	156	S52246	transposable retro
13	42	47.2	243	T15225	hypothetical prote
14	42	47.2	324	T10685	DNA-binding protei
15	42	47.2	573	T21355	hypothetical prote
16	42	47.2	820	C75284	phenylalanyl-tRNA
17	42	47.2	5032	A35041	ryanodine receptor
18	41	46.1	40	B61320	plastocyanin b - L
19	41	46.1	99	S00210	plastocyanin b - L
20	41	46.1	155	S38255	plastocyanin b pre
21	41	46.1	168	S58208	plastocyanin b pre
22	41	46.1	226	E70809	hypothetical prote
23	41	46.1	281	JC6528	31K major protein, Esc
24	41	46.1	286	I80309	ribose-phosphate p
25	41	46.1	318	S37225	p-aminobenzoate sy
26	41	46.1	348	F70398	probable ribose-ph
27	41	46.1	355	S45804	lipid A disacchari
28	41	46.1	360	C64628	escN [Imported]
29	41	46.1	446	D86046	

30 catalase (EC 1.11.1.1)  
31 ArgE/DapE/AcyI fam  
32 probable regulator  
33 outer membrane ush  
34 elastic titin - hu  
35 plastocyanin - fie  
36 hypotheical prote  
37 hypotheical prote  
38 homeotic protein s  
39 hypotheical prote  
40 conserved hypochet  
41 GumC protein - Xan  
42 conserved hypochet  
43 thiorodoxin reduct  
44 cell surface glyco  
45 myosin Myok - Dict

ALIGNMENTS

RESULT 1

T50454  
probable rho1 GDP-GTP exchange protein [imported] - fission yeast (Schizosaccharomy  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T50454  
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Simmonds, M.; Churcher, C.M.  
submitted to the EMBL Data Library, November 1999  
A:Reference number: Z25030  
A:Accession: T50454  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1158 <MCD>  
A:Cross-references: EMBL:AL132828; PIDN:CA960236.1; GSPDB:GN000066; SPDB:SPAC1006.06  
A:Experimental source: strain 972h(-); cosmid cl006  
C:Genetics:  
A:Gene: SPDB:SPAC1006.06  
A:Map position: 1  
A:Introns: 835/1, 975/3  
C:Superfamily: CDC24 homology

Query Match 52.8%; Score 47; DB 2; Length 1158;  
Best Local Similarity 55.6%; Pred. No. 20;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEEFSVSUGE 18  
||||: ||| | ||  
DB 1122 EGGEILYSTETPIPFSSGE 1139

RESULT 2

S21102  
lectin - eel (Conger myriaster)  
C:Species: Conger myriaster  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 26-Aug-1999  
C:Accession: S21102  
R:Muramoto, K.; Kamiya, H.  
Biochim. Biophys. Acta 1116, 129-136, 1992  
A:Title: The amino-acid sequence of a lectin from conger eel, Conger myriaster, sk  
A:Reference number: S21102; MUID:92256465  
A:Accession: S21102  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-135 <MUR>  
C:Superfamily: beta-galactoside-binding lectin

Query Match 50.6%; Score 45; DB 2; Length 135;  
Best Local Similarity 52.9%; Pred. No. 4.1;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVMGE 18  
 Db 20 GGFNNSPQFVSVMGE 36

RESULT 3  
 D49339  
 hypothetical protein (bbsC 5' region) - Enterococcus faecalis  
 C:Species: Enterococcus faecalis  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 20-Oct-2000  
 C:Accession: D49339  
 R:Bensing, B.A.; Dunny, G.M.  
 J. Bacteriol. 175, 7421-7429, 1993  
 A:Title: Cloning and molecular analysis of genes affecting expression of binding substance  
 A:Reference number: A49339; MUID:94042918  
 A:Accession: D49339  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-253 <BEN>  
 A:Cross-references: GB:L23802; NID:9388106; PIDN:AAC36854.1; PID:g388110  
 C:Genetics:  
 A:Start codon: GTG  
 C:Superfamily: 3-dehydroquinase dehydratase; 3-dehydroquinase dehydratase homology  
 F:19-239/Domain: 3-dehydroquinase dehydratase homology <DQD>

Query Match 50.6%; Score 45; DB 1; Length 253;  
 Best Local Similarity 58.3%; Pred. No. 8.1;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGELSYSEEF 12  
 Db 86 EGGEAFSENY 97

RESULT 4  
 H75357  
 tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase - Deinococcus radiodurans  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: H75357  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: H75357  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-413 <WHI>  
 A:Cross-references: GB:AE002017; GB:AE000513; NID:g6459527; PIDN:AAF11314.1; PID:g645953  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRI759  
 A:Map position: 1  
 C:Superfamily: probable membrane protein YDL033c

Query Match 50.6%; Score 45; DB 2; Length 413;  
 Best Local Similarity 44.4%; Pred. No. 14;  
 Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 EGELSYSEEFVSVMGE 18  
 Db 369 EGFELEAFEPQFAPQ 386

RESULT 5  
 CUPX  
 plastocyanin a precursor [validated] - Lombardy poplar  
 C:Species: Populus nigra var. italica (Lombardy poplar)  
 C>Date: 31-May-1980 #sequence\_revision 31-Oct-1997 #text\_change 15-Sep-2000

C:Accession: S58209; A00309  
 R:Reichert, J.; Jenzelewski, V.; Haehnel, W.  
 submitted to the EMBL Data Library, July 1995  
 A:Description: Kinetic studies of recombinant poplar plastocyanins.  
 A:Reference number: S58208  
 A:Accession: S58209  
 A:Molecule type: mRNA  
 A:Residues: 1-168 <REN>  
 A:Cross-references: EMBL:Z50185; NID:929812; PIDN:CAA90564.1; PID:g929813  
 A:Experimental source: var. italica  
 R:Ambler, R.  
 unpublished results, cited by Freeman, H.C., J. Proc. Royal Soc. N.S. Wales 112, 45-  
 A:Reference number: A94471  
 A:Accession: A00309  
 A:Molecule type: protein  
 A:Residues: 70-127, 'ZZB', 131-168 <AMB>  
 R:Guss, J.M.; Freeman, H.C.  
 submitted to the Brookhaven Protein Data Bank, March 1992  
 A:Reference number: A51342; PDB:1PFC  
 A:Contents: annotation; X-ray crystallography, 1.33 angstroms, residues 70-168  
 R:Guss, J.M.; Freeman, H.C.  
 submitted to the Brookhaven Protein Data Bank, September 1986  
 A:Reference number: A50737; PDB:5PCY  
 A:Contents: annotation; X-ray crystallography, 1.80 angstroms, residues 70-168  
 R:Guss, J.M.; Harrowell, P.R.; Murata, M.; Norris, V.A.; Freeman, H.C.  
 J. Mol. Biol. 192, 361-387, 1986  
 A:Title: Crystal structure analyses of reduced (Cu(I)) poplar plastocyanin at six pH  
 A:Reference number: A58637; MUID:87169729  
 A:Contents: annotation; X-ray crystallography, 1.80 angstroms  
 R:Guss, J.M.; Freeman, H.C.  
 J. Mol. Biol. 169, 521-563, 1983  
 A:Title: Structure of oxidized poplar plastocyanin at 1.6 Angstroms resolution.  
 A:Reference number: A58639; MUID:84010876  
 A:Contents: annotation; X-ray crystallography, 1.60 angstroms  
 R:Colman, P.M.; Freeman, H.C.; Guss, J.M.; Murata, M.; Norris, V.A.; Ramshaw, J.A.M.  
 Nature 272, 319-324, 1978  
 A:Title: X-ray crystal structure analysis of plastocyanin at 2.7 angstrom resolution  
 A:Reference number: A93194  
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms  
 C:Comment: Plastocyanin is found loosely bound to the inner thylakoid membrane surfa  
 C:Genetics:  
 A:Gene: petE  
 A:Genome: nuclear  
 C:Function:  
 A:Description: accepts electrons from cytochrome f and donates electrons to photosys  
 C:Superfamily: plastocyanin  
 C:Keywords: chloroplast; copper; electron transfer; membrane-associated protein; met  
 F:1-69/Domain: transit peptide (chloroplast) #status predicted <TRP>  
 F:70-168/Product: plastocyanin a #status experimental <MAT>  
 F:106,153,156,161/Binding site: copper (His, Cys, His, Met) (type 1) #status experim

Query Match 48.3%; Score 43; DB 1; Length 168;  
 Best Local Similarity 44.4%; Pred. No. 11;  
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 EGELSYSEEFVSVMGE 18  
 Db 77 DDGSLAFVPSEFSPGE 94

RESULT 6  
 QQAG6T  
 hypothetical protein 6 - Agrobacterium tumefaciens plasmids  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 16-Jul-1999  
 C:Accession: A04498; S28691  
 R:Gielen, J.; De Beuckeleer, M.; Seurinck, J.; Deboeck, F.; De Greve, H.; Lemmers, M  
 EMBO J. 3, 835-846, 1984  
 A:Title: The complete nucleotide sequence of the TL-DNA of the Agrobacterium tumefac  
 A:Reference number: A91001; MUID:84207942  
 A:Accession: A04498  
 A:Molecule type: DNA

A;Residues: 1-191 <GIE>  
A;Cross-references: GB:X00493; GB:J05108; GB:X00282; NID:g39062; PIDN:CAA25171.1; PID:g39071  
A;Experimental source: plasmid pTiAch5  
R;Barker, R.F.; Idler, K.B.; Thompson, D.V.; Kemp, J.D.  
Plant Mol. Biol. 2, 335-350, 1983  
A;Title: Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine plasmid  
A;Reference number: S28683  
A;Accession: T28691  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-191 <BAR>  
A;Cross-references: EMBL:X00493; NID:g39062; PIDN:CAA25171.1; PID:g39071  
A;Experimental source: plasmid pTi15955  
C;Genetics:  
A;Genome: plasmid  
C;Superfamily: T-6b protein  
C;Keywords: crown gall tumor

Query Match 48.3%; Score 43; DB 1; Length 191;  
Best Local Similarity 46.7%; Pred. No. 13;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVS 15  
:|:|::||:|:|  
Db 111 DGGRIYSKNEYSS 125

RESULT 7  
D84365  
electron transfer flavoprotein subunit alpha [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: D84365  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo, J.; Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L...  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483  
A;Accession: D84365  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-322 <STO>  
A;Cross-references: GB:AE004437; NID:g10581567; PIDN:AAG20288.1; GSPDB:GN00138  
C;Genetics:  
A;Gene: etfa  
C;Superfamily: electron transfer flavoprotein alpha chain

Query Match 48.3%; Score 43; DB 2; Length 322;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVS 17  
:|:|::||:|:|  
Db 192 GGDVDIADAEFLVSG 207

RESULT 8  
T09272  
probable tail component protein 515 - Streptococcus thermophilus phage Sfil9  
C;Species: Streptococcus thermophilus phage Sfil9  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: T09272  
R;Desiere, F.; Lucchini, S.; Brussow, H.  
Virology 241, 345-356, 1998  
A;Title: Evolution of Streptococcus thermophilus bacteriophage genomes by modular exchange  
A;Reference number: Z16607; MUID:98160788  
A;Accession: T09272  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA

A;Residues: 1-515 <DES>  
A;Cross-references: EMBL:AF032122; NID:g5523999; PIDN:AAC39296.1; PID:g2935690  
C;Superfamily: Streptococcus phage Sfil9 probable tail protein 515  
C;Keywords: tail protein

Query Match 48.3%; Score 43; DB 2; Length 515;  
Best Local Similarity 53.3%; Pred. No. 37;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 ELSYSEEFVS 18  
:|:|::||:|:|  
Db 475 ELTYLSEPFSGTGE 489

RESULT 9  
T09203  
probable tail component protein 515 - Streptococcus thermophilus phage Sfil21  
C;Species: Streptococcus thermophilus phage Sfil21  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: T09203  
R;Desiere, F.; Lucchini, S.; Brussow, H.  
Virology 241, 345-356, 1998  
A;Title: Evolution of Streptococcus thermophilus bacteriophage genomes by modular exchange  
A;Reference number: Z16607; MUID:98160788  
A;Accession: T09203  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-515 <DES>  
A;Cross-references: EMBL:AF032121; NID:g5524032; PIDN:AAC39282.1; PID:g2935675  
C;Superfamily: Streptococcus phage Sfil9 probable tail protein 515  
C;Keywords: tail protein

Query Match 48.3%; Score 43; DB 2; Length 515;  
Best Local Similarity 53.3%; Pred. No. 37;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 ELSYSEEFVS 18  
:|:|::||:|:|  
Db 475 ELTYLSEPFSGTGE 489

RESULT 10  
G70327  
isoleucine--tRNA ligase (EC 6.1.1.5) - Aquifex aeolicus  
C;Species: Aquifex aeolicus  
C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Jul-1999  
C;Accession: G70327  
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.  
V.

Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A;Reference number: A70300; MUID:98196666  
A;Accession: G70327  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-956 <AQE>  
A;Cross-references: GB:AE000683; NID:g2982996; PIDN:AAC06614.1; PID:g2983002; GB:AE000683  
A;Experimental source: strain VFS  
C;Genetics:  
A;Gene: ileS  
C;Superfamily: isoleucine--tRNA ligase  
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 48.3%; Score 43; DB 2; Length 956;  
Best Local Similarity 52.9%; Pred. No. 73;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEFVS 17  
:|:|::||:|:|  
Db 896 EGGVQTEGELPKVKG 912

```

RESULT 11
S20969
Na+/Ca2+, K+-exchanging protein - bovine
N:Alternate names: Na+/Ca2+, K+ antiporter; Na/Ca, K-exchanger
C:Species: Bos primigenius taurus (cattle)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C:Accession: S20969
R:Reilaender, H.; Achilles, A.; Friedel, U.; Maul, G.; Lottspeich, F.; Cook, N.J.
EMBO J. 11, 1689-1695, 1992
A:Title: Primary structure and functional expression of the Na/Ca,K-exchanger from bovine
A:Reference number: S20969; MUID:92258377
A:Accession: S20969
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1199 <RE>
A:Cross-references: GB:X66481; NID:9505578; PIDN:CAA47108.1; PID:9505579

Query Match 48.3%; Score 43; DB 2; Length 1199;
Best Local Similarity 44.4%; Pred. No. 93;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 EGGELSYSEEEFSVSUGE 18
      |||||  I:I  :  ||
Db 805 EGGEVKGDEGEIQAGE 822

RESULT 12
S52246
transposable retroelement homolog - perennial teosinte
N:Alternate names: ZLR57 protein, long repetitive sequence protein ZLRS 7
C:Species: Zea diploperennis (perennial teosinte)
C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 07-Dec-1999
C:Accession: JC4641; S52246
R:Monfort, A.; Vicent, C.M.; Raz, R.; Puigdomenech, P.; Martinez-Izquierdo, J.A.
DNA Res. 2, 255-261, 1995
A:Title: Molecular analysis of a putative transposable retroelement from the Zea genus
A:Reference number: JC4641; MUID:97021439
A:Accession: JC4641
A:Molecule type: DNA
A:Residues: 1-156 <MON>
A:Cross-references: EMBL:X82087; NID:9609287; PIDN:CAA57619.1; PID:9609288
F:123-153/Region: DNA binding #status predicted

Query Match 47.2%; Score 42; DB 2; Length 156;
Best Local Similarity 72.7%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 GELSYSEEEFS 13
      I:  |||||
Db 19 GDSEYSEEEFS 29

RESULT 13
T15225
hypothetical protein C55C2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15225
R:Goela, D.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid C55C2.
A:Reference number: Z18311
A:Accession: T15225
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-243 <GOE>
A:Cross-references: EMBL:AF003144; NID:92088754; PID:92088759; PIDN:AAB54197.1; GSPDB:GN
A:Experimental source: strain Bristol N2; clone C55C2
C:Genetics:

```

```

A:Gene: CESP:C55C2.1
A:Map position: 1
A:Introns: 8/3; 41/1; 53/3; 140/1; 164/2; 211/1

Query Match 47.2%; Score 42; DB 2; Length 243;
Best Local Similarity 41.2%; Pred. No. 24;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEEFSVSUGE 18
      ||:|:|:|:|:|:|
Db 168 GGDLLHFFKKKFNFSAGE 184

RESULT 14
T10685
DNA-binding protein WRKY3 homolog F3L17.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T10685
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16652
A:Accession: T10685
A:Molecule type: DNA
A:Residues: 1-324 <BEV>
A:Cross-references: EMBL:AL080283; GSPDB:GN00062; ATSP:F3L17.120
A:Experimental source: cultivar Columbia; BAC clone F3L17
C:Genetics:
A:Gene: ATSP:F3L17.120
A:Map position: 4
A:Introns: 223/2; 265/2

Query Match 47.2%; Score 42; DB 2; Length 324;
Best Local Similarity 50.8%; Pred. No. 33;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 EGGELSYSEEEFSVSU 16
      :  ||  :|:|:|:|
Db 135 KSABLEFSKENFSVSL 150

RESULT 15
T21355
hypothetical protein F25H2.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21355
R:Wilkinson, J.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19411
A:Accession: T21355
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-573 <WIL>
A:Cross-references: EMBL:Z79754; PIDN:CAB02094.2; GSPDB:GN00019; CESP:F25H2.6
A:Experimental source: clone F25H2
C:Genetics:
A:Gene: CESP:F25H2.6
A:Map position: 1
A:Introns: 5/2; 60/3; 100/3; 205/3; 268/3; 320/3; 353/2; 487/3; 532/1

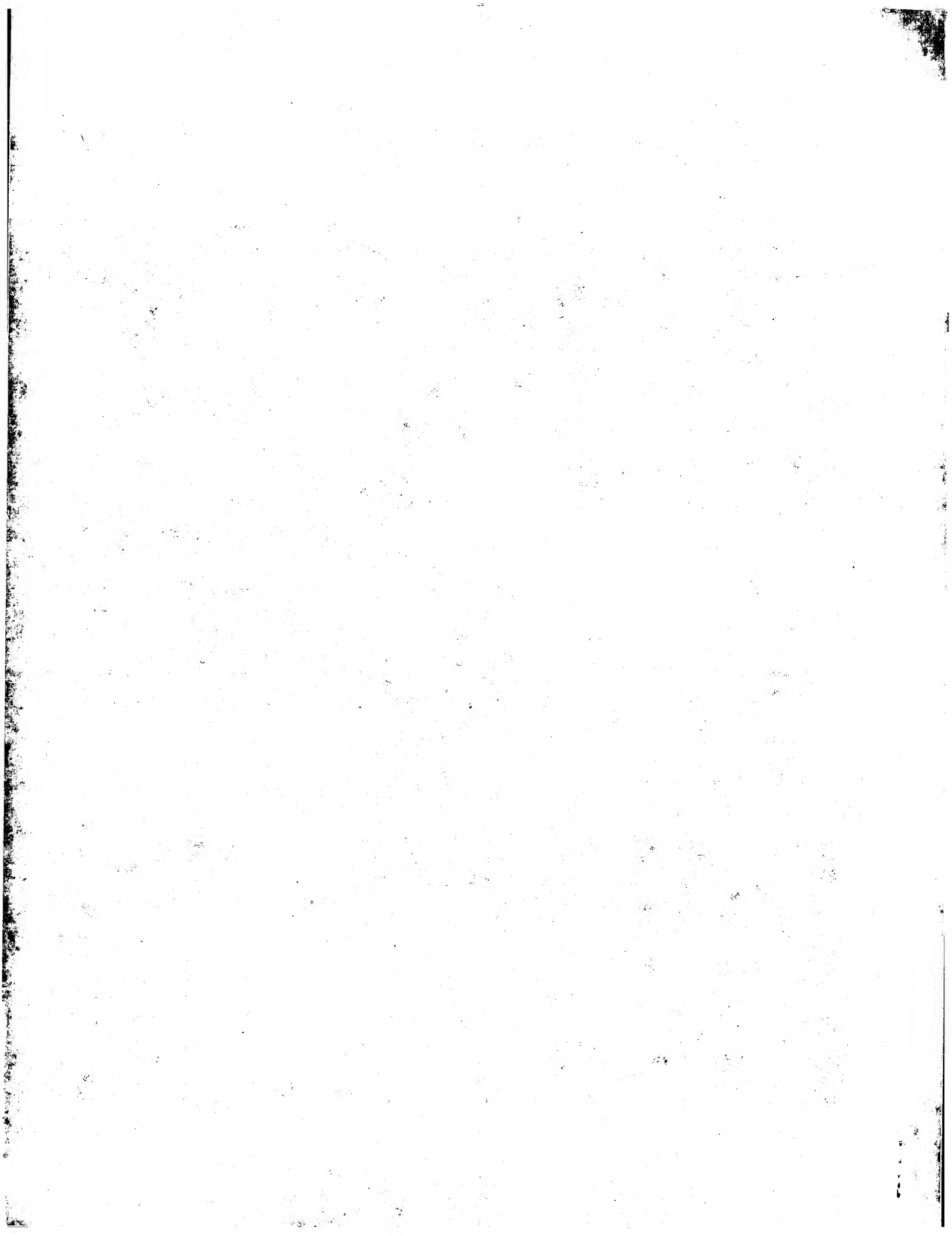
Query Match 47.2%; Score 42; DB 2; Length 573;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 YSEEEFSVSUGE 18
      :  |||||:|:|
Db 64 FDENEFSISMGE 75

```

Search completed: February 12, 2002, 12:34:36  
Job time: 554 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:22 ; Search time 106.12 Seconds  
(without alignments)  
3.817 Million cell updates/sec

Title: US-09-485-571-18  
Perfect score: 89  
Sequence: 1 EGGELSYSEEEFSVSGE 18  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_5/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_5/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_5/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_5/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_5/ptodata/2/iaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_5/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	46	51.7	18	1 US-08-182-483A-25	Sequence 25, Appl
2	46	51.7	18	1 US-08-243-879A-24	Sequence 24, Appl
3	46	51.7	18	1 US-08-499-523-45	Sequence 45, Appl
4	46	51.7	18	4 US-09-128-345-45	Sequence 45, Appl
5	44	49.4	307	1 US-08-713-828-1	Sequence 1, Appl
6	44	49.4	307	2 US-08-919-627-1	Sequence 1, Appl
7	44	49.4	307	2 US-09-096-245-1	Sequence 1, Appl
8	43	48.3	18	1 US-08-499-523-63	Sequence 63, Appl
9	43	48.3	18	1 US-08-499-523-67	Sequence 67, Appl
10	43	48.3	18	2 US-08-752-852A-230	Sequence 230, App
11	43	48.3	18	4 US-09-128-345-63	Sequence 63, Appl
12	43	48.3	18	4 US-09-128-345-67	Sequence 67, Appl
13	42	47.2	18	1 US-08-182-483A-29	Sequence 29, Appl
14	42	47.2	18	1 US-08-243-879A-28	Sequence 28, Appl
15	42	47.2	18	1 US-08-499-523-49	Sequence 49, Appl
16	42	47.2	18	4 US-09-128-345-49	Sequence 49, Appl
17	41	46.1	18	1 US-08-182-483A-28	Sequence 28, Appl
18	41	46.1	18	1 US-08-243-879A-27	Sequence 27, Appl
19	41	46.1	18	1 US-08-499-523-48	Sequence 48, Appl
20	41	46.1	18	1 US-08-499-523-53	Sequence 53, Appl
21	41	46.1	18	1 US-08-499-523-58	Sequence 58, Appl
22	41	46.1	18	1 US-08-499-523-58	Sequence 58, Appl
23	41	46.1	18	4 US-09-128-345-48	Sequence 48, Appl
24	41	46.1	18	4 US-09-128-345-53	Sequence 53, Appl
25	41	46.1	18	4 US-09-128-345-58	Sequence 58, Appl
26	41	46.1	18	4 US-09-128-345-65	Sequence 65, Appl
27	41	46.1	291	4 US-09-201-641-4	Sequence 4, Appl

28	40	44.9	18	1 US-08-499-523-54	Sequence 54, Appl
29	40	44.9	18	1 US-08-499-523-57	Sequence 57, Appl
30	40	44.9	18	1 US-08-499-523-59	Sequence 59, Appl
31	40	44.9	18	1 US-08-499-523-62	Sequence 62, Appl
32	40	44.9	18	2 US-08-752-852A-227	Sequence 227, App
33	40	44.9	18	2 US-08-752-852A-228	Sequence 228, App
34	40	44.9	18	2 US-08-752-852A-229	Sequence 229, App
35	40	44.9	18	3 US-08-752-853-9	Sequence 9, Appl
36	40	44.9	18	4 US-09-128-345-54	Sequence 54, Appl
37	40	44.9	18	4 US-09-128-345-57	Sequence 57, Appl
38	40	44.9	18	4 US-09-128-345-59	Sequence 59, Appl
39	40	44.9	18	4 US-09-128-345-62	Sequence 62, Appl
40	40	44.9	549	2 US-08-500-635A-12	Sequence 12, Appl
41	40	44.9	549	4 US-09-167-151-12	Sequence 12, Appl
42	39	43.8	17	2 US-08-752-852A-77	Sequence 77, Appl
43	39	43.8	18	1 US-08-095-769A-1	Sequence 1, Appl
44	39	43.8	18	1 US-08-182-483A-2	Sequence 2, Appl
45	39	43.8	18	1 US-08-182-483A-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1  
US-08-182-483A-25  
; Sequence 25, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KORYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,483A  
; FILING DATE: 13-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-182-483A-25

Query Match 51.7%; Score 46; DB 1; Length 18;  
Best Local Similarity 56.2%; Pred. No. 0.2;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 GGELSYSEEEFSVSG 17  
||| |:::||||  
Db 2 GGRLCYSRRKWCVSVG 17

```

RESULT 2
US-08-243-879A-24
; Sequence 24, Application US/08243879A
; Patent No. 5708145
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; TITLE OF INVENTION: A NEW PROTEGRIN
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington, DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/243,879A
; FILING DATE: 17-MAY-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-243-879A-24

```

```

Query Match 51.7%; Score 46; DB 1; Length 18;
Best Local Similarity 56.2%; Pred. No. 0.2;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 GGELSYSEEFVSVG 17
||| ||| :: ||||
Db 2 GGRCLYRKKWCVSIG 17

```

```

RESULT 3
US-08-499-523-45
; Sequence 45, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-499-523-45

```

```

Query Match 51.7%; Score 46; DB 1; Length 18;
Best Local Similarity 56.2%; Pred. No. 0.2;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 GGELSYSEEFVSVG 17
||| ||| :: ||||
Db 2 GGRCLYRKKWCVSIG 17

```

```

RESULT 4
US-09-128-345-45
; Sequence 45, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-128-345-45

```

Query Match 51.7%; Score 46; DB 4; Length 18;  
Best Local Similarity 56.2%; Pred. No. 0.2;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEEFSVSG 17  
DB 2 GGRLCYSRKKWCVSG 17

## RESULT 5

US-08-713-828-1  
; Sequence 1, Application US/08713828  
; Patent No. 5683910  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHORYLASE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/713,828  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0068 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 307 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE: Consensus  
; US-08-713-828-1

Query Match 49.4%; Score 44; DB 1; Length 307;  
Best Local Similarity 90.0%; Pred. No. 9.8;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LSYSEEEFSV 14  
DB 189 LSYSEEEFDV 198

## RESULT 6

US-08-919-627-1  
; Sequence 1, Application US/08919627  
; Patent No. 5833981  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga

; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHORYLASE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/919,627  
; FILING DATE: August 28, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/713,828  
; FILING DATE: September 13, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0068-1 DIV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 307 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE: Consensus  
; US-08-919-627-1

Query Match 49.4%; Score 44; DB 2; Length 307;  
Best Local Similarity 90.0%; Pred. No. 9.8;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LSYSEEEFSV 14  
DB 189 LSYSEEEFDV 198

## RESULT 7

US-09-096-245-1  
; Sequence 1, Application US/09096245  
; Patent No. 5977320  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHORYLASE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/096,245  
;; FILING DATE:  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: 08/713,828  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Billings, Lucy J.  
;; REGISTRATION NUMBER: 36,749  
;; REFERENCE/DOCKET NUMBER: PF-0068 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-855-0555  
;; TELEFAX: 415-845-4166  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 307 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; IMMEDIATE SOURCE:  
;; LIBRARY:  
;; CLONE: Consensus  
US-09-096-245-1

Query Match 49.4%; Score 44; DB 2; Length 307;  
Best Local Similarity 90.0%; Pred. No. 9.8;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LSYSEEFVS 14  
|||||||  
DB 189 LSYSEEFV 198

## RESULT 8

US-08-499-523-63  
; Sequence 63, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELEPHONE: (202) 887-0763  
; TELEFAX: (202) 887-1500  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid

;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: group(6, 8, 13, 15)  
;; OTHER INFORMATION: /note= "X is a hydrophobic, a  
;; OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-63

Query Match 48.3%; Score 43; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.6;  
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVS 17  
|||||  
DB 2 GGRLXXRRRFXV 17  
|||||

## RESULT 9

US-08-499-523-67  
; Sequence 67, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 67:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 8, 13, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-67

Query Match 48.3%; Score 43; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.6;  
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVS 17  
|||||

Db 2 GGRLXXRRRFXVXVG 17

## RESULT 10

US-08-752-852A-230  
; Sequence 230, Application US/08752852A  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrer, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 230:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; US-08-752-852A-230

Query Match 48.3%; Score 43; DB 2; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.6;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGELSYSEEEFVSXVG 17

Db 2 GGRLCYARRRFAVCVG 17

## RESULT 11

US-09-128-345-63  
; Sequence 63, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S. L.  
; APPLICANT: KORYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-0054-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: Group(6, 8, 13, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
; US-09-128-345-63

Query Match 48.3%; Score 43; DB 4; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.6;  
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 GGELSYSEEEFVSXVG 17

Db 2 GGRLXXRRRFXVXVG 17

## RESULT 12

US-09-128-345-67  
; Sequence 67, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S. L.  
; APPLICANT: KORYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 8, 13, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
US-09-128-345-67  
OTHER INFORMATION: small, or a large polar amino acid"

Query Match 48.3%; Score 43; DB 4; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.6;  
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVG 17  
||| | | | | |  
Db 2 GGLXYRRRFXVVG 17

RESULT 13  
US-08-182-483A-29  
; Sequence 29, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/182.483A  
FILING DATE: 13-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 10

Query Match 48.3%; Score 43; DB 4; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.6;  
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVG 17  
||| | | | | |  
Db 2 GGLXYRRRFXVVG 17

..

OTHER INFORMATION: /product= "homoarginine(Har)"  
US-08-182-483A-29  
Query Match 47.2%; Score 42; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.88;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVG 17  
||| | | | | |  
Db 2 GGLXYRRRFXVVG 17

RESULT 14  
US-08-243-879A-28  
; Sequence 28, Application US/08243879A  
; Patent No. 5708145  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: A NEW PROTEGRIN  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington, DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/243,879A  
FILING DATE: 17-MAY-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 10  
OTHER INFORMATION: /note= "This position is Har."  
US-08-243-879A-28

Query Match 47.2%; Score 42; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.88;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGELSYSEEFVSVG 17  
||| | | | | |  
Db 2 GGLXYRRRFXVVG 17

RESULT 15  
US-08-499-523-49  
; Sequence 49, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.

APPLICANT: HAWIG, SYLVIA S.L.  
APPLICANT: KORAYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,523  
FILING DATE: 07-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 10  
OTHER INFORMATION: /product= "homoarginine(Har)"  
US-08-499-523-49

Query Match 47.2%; Score 42; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 0.88;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGELSYSEEEFSVSG 17  
Db 2 GGRVCYGRXRFVCGV 17

Search completed: February 12, 2002, 12:32:22  
Job time: 450 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:29 ; Search time 242.57 seconds  
(without alignments)  
5.497 Million cell updates/sec

Title: US-09-485-571-18

Perfect score: 89

Sequence: 1 EGGELSYSEEFVSUGE 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	89	100.0	18	20 AAW99405	Protegrin derivati
2	59	66.3	18	20 AAW99403	Protegrin derivati
3	59	66.3	18	21 AA193616	Peptide which may
4	51	57.3	18	20 AAW99412	Protegrin derivati
5	51	57.3	18	21 AA193177	Protegrin-like pep
6	51	57.3	18	21 AA193179	Protegrin-like pep
7	51	57.3	18	21 AA193615	Peptide which may
8	46	51.7	18	16 AAW78773	Protegrin peptide
9	45	50.6	41	21 AAB30450	Amino acid sequenc
10	44	49.4	307	18 AAW34892	Novel human phosph
11	44	49.4	307	20 AAW76803	Human phosphorilas

12	44	49.4	307	21	AA152303	Novel human phosph
13	44	49.4	414	20	AA127161	Human DRK1 protei
14	44	49.4	414	22	AA165625	Novel protein kina
15	43	48.3	18	18	AA136429	Antimicrobial prot
16	43	48.3	18	18	AA109084	Antimicrobial prot
17	43	48.3	18	18	AA109085	Cationic, antimicr
18	42	47.2	18	16	AA178777	Protegrin peptide
19	42	47.2	18	18	AA181511	Cationic, antimicr
20	42	47.2	18	18	AA181512	Cationic, antimicr
21	42	47.2	319	21	AA111603	Arabidopsis thalia
22	42	47.2	319	21	AA142800	Arabidopsis thalia
23	42	47.2	324	21	AA108140	Arabidopsis thalia
24	42	47.2	324	21	AA111602	Arabidopsis thalia
25	42	47.2	324	21	AA142799	Arabidopsis thalia
26	42	47.2	324	22	AA101904	Arabidopsis thalia
27	42	47.2	324	22	AA101921	Arabidopsis thalia
28	42	47.2	329	21	AA108139	Arabidopsis thalia
29	42	47.2	350	21	AA142798	Arabidopsis thalia
30	42	47.2	353	21	AA111601	Arabidopsis thalia
31	42	47.2	356	21	AA108138	Arabidopsis thalia
32	41	46.1	5072	12	AA111510	Ryanodine receptor
33	41	46.1	15	20	AA109411	Protegrin derivati
34	41	46.1	18	16	AA178776	Protegrin peptide
35	41	46.1	255	21	AA132322	Soybean beta-carot
36	41	46.1	291	21	AA190227	Marigold beta-hydr
37	41	46.1	314	21	AA132321	Soybean beta-carot
38	40	44.9	18	18	AA136426	Antimicrobial prot
39	40	44.9	18	18	AA136427	Antimicrobial prot
40	40	44.9	18	18	AA136428	Antimicrobial prot
41	40	44.9	18	18	AA135585	Antimicrobial pept
42	40	44.9	18	18	AA181513	Cationic, antimicr
43	40	44.9	18	18	AA109081	Cationic, antimicr
44	40	44.9	18	18	AA109082	Cationic, antimicr
45	40	44.9	18	18	AA109083	Cationic, antimicr

ALIGNMENTS

RESULT 1

AA199405  
ID AAW99405 standard; peptide; 18 AA.

AC AAW99405;

DT 08-JUN-1999 (first entry)

DE Protegrin derivative peptide SMI739.

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

OS Synthetic.

PN WO9907728-A2.

XX 18-FEB-1999.

PD 06-AUG-1998; 98WO-FR01757.

XX 12-AUG-1997; 97FR-0010297.

PA (SYNT-) SYNT:EM SA.

PI Calas E, Chavanieu A, Grassy G, Kaczorek M;

XX WPI; 1999-190034/16.

DR Derivatives of antibiotic peptides lacking disulfide bridges - used  
XX as carriers to deliver active agents into cells

PS Claim 7; Page 28; 37pp; French.

CC This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX SQ Sequence 18 AA;

Query Match 100.0%; Score 89; DB 20; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-08; Mismatches 0; Indels 0; Gaps 0;  
 Matches 18; Conservative 0;

QY 1 EGGELSYSEEFSSVSGE 18  
 |||||

Db 1 eggelsyseefssvsg 18

RESULT 2

AAW99403  
 ID AAW99403 standard; peptide; 18 AA.

AC AAW99403;

DT 08-JUN-1999 (first entry)

DE Protegrin derivative peptide SM1738.

KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

XX Synthetic.

OS WO9907728-A2.

PN 18-FEB-1999.

PD 06-AUG-1998; 98WO-FR01757.

PF 12-AUG-1997; 97FR-0010297.

PR (SYNT-) SYNT:EM SA.

PA Calas B, Chavanieu A, Grassy G, Kaczorek M;  
 PI WPI; 1999-190034/16.

DR Derivatives of antibiotic peptides lacking disulfide bridges - used  
 as carriers to deliver active agents into cells

PT Claim 7; Page 28; 37pp; French.

PS This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX SQ Sequence 18 AA;

Query Match 66.3%; Score 59; DB 20; Length 18;  
 Best Local Similarity 75.0%; Pred. No. 0.0028;  
 Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEFSSVSG 17  
 |||||

Db 2 ggrrlsysrrrrfsvsg 17

RESULT 3

AA93616  
 ID AAY93616 standard; peptide; 18 AA.

AC AAY93616;

DT 25-SEP-2000 (first entry)

DE Peptide which may be linked to anticancer agents.

KW Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 KW cancer.

XX Unidentified.

PN WO200032237-A1.

XX 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02939.

XX 30-NOV-1998; 98FR-0015073.

XX (SYNT-) SYNT:EM SA.

XX Tamsamani J, Kaczorek M, Colin De Verdiere A;  
 XX WPI; 2000-412166/35.

XX New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells  
 PS Disclosure; Page 8; 34pp; French.

CC The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.

XX SQ Sequence 18 AA;

Query Match 66.3%; Score 59; DB 21; Length 18;  
 Best Local Similarity 75.0%; Pred. No. 0.0028;  
 Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGELSYSEEFSSVSG 17  
 |||||

Db 2 ggrrlsysrrrrfsvsg 17

RESULT 4

AAW99412  
ID AAW99412 standard; peptide; 18 AA.  
XX  
AC AAW99412;  
XX  
DT 08-JUN-1999 (first entry)  
XX  
DE Protegrin derivative peptide SM2196.  
XX  
KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.  
XX  
OS Synthetic.  
XX  
PN WO9907728-A2.  
XX  
PD 18-FEB-1999.  
XX  
PF 06-AUG-1998; 98WO-FR01757.  
XX  
PR 12-AUG-1997; 97FR-0010297.  
XX  
PA (SYNT-) SYNT:EM SA.  
XX  
PI Calas B, Chavanleu A, Grassy G, Kaczorek M;  
XX WPI; 1999-190034/16.  
XX  
DR Derivatives of antibiotic peptides lacking disulfide bridges - used  
XX as carriers to deliver active agents into cells  
XX  
PS Claim 7; Page 28; 37pp; French.  
XX  
CC This peptide represents a linear derivative of the protegrin family of  
CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
CC antibiotic family which contain a beta-sheet secondary structure linked  
CC by disulphide bridges. The new derivatives are linear and lack the  
CC disulphide bridge. The novel derivatives are used to deliver active  
CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
CC including crossing the blood-brain barrier.  
XX  
SQ Sequence 18 AA;

Query Match 57.3%; Score 51; DB 20; Length 18;  
Best Local Similarity 62.5%; Pred. No. 0.06;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGELSYSEEFVSVG 17  
|| |||| || ||  
Db 2 ggrlsysrrfststg 17

RESULT 5  
AA93177  
ID AAY93177 standard; peptide; 18 AA.  
XX  
AC AAY93177;  
XX  
DT 06-DEC-2000 (first entry)  
XX  
DE Protegrin-like peptide antibiotic Doxo-SynBI.  
XX  
KW Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
KW cancer; Parkinson's disease; depression; pain; meningitis.

XX Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1 /note- "linked to doxorubicin via a succinate  
FT (-CO-(CH2)2-CO-) linker; optionally linked  
FT to benzylpenicillin by a glycoamide linker"  
XX  
PN WO200032236-A1.  
XX  
PD 08-JUN-2000.  
XX  
PF 26-NOV-1999; 99WO-FR02938.  
XX  
PR 30-NOV-1998; 98FR-0015074.  
XX (SYNT-) SYNT:EM SA.  
XX  
PI Clair P, Kaczorek M, Tamsamani J;  
XX WPI; 2000-422871/36.  
XX  
XX Use of linear peptides as vectors for active ingredients, useful for  
PT diagnosis and treatment of central nervous system diseases, can  
PT transport agents passively across the blood-brain barrier -  
XX  
PS Example I; Page 13; 54pp; French.  
XX  
CC The invention relates to the use of linear peptides, coupled to an active  
CC agent, to prepare a composition able to cross the blood-brain barrier  
CC for diagnosis or treatment of disorders localised in the central nervous  
CC system. The linear peptide preferably has the formula: (a) X1-X16;  
CC (b): BXBXBXBXBXBXBXBX; or (c) BXXBXBXBXBXBXBXBX, where: each of X1-X16  
CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
CC be trp; each B is aa containing a side chain that includes a basic group;  
CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
CC Peptides able to cross the BBB include protegrins, Antennapedia,  
CC tachyplesins, transportin, etc. Of these several families have cytolytic  
CC effects and are termed peptide antibiotics. They fall into 3 main  
CC categories based on their structure: (i) peptides with alpha-helices,  
CC e.g. ceropins and maganins; (ii) peptides with disulphide bond-linked  
CC beta-sheets, e.g. protegrin, tachyplesins, defensins; (iii) peptides  
CC with no major structure but containing bends due to the presence of  
CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
CC fall into the peptide antibiotic categories defined above: (a)-peptides  
CC are based on the Antennapedia family peptides; (b)-peptides are based on  
CC protegrins; and (c)-peptides are based on tachyplesins. This sequence  
CC represents a synthetic linear peptide designed on peptides able to cross  
CC the BBB and is conjugated to a doxorubicin molecule by a succinate  
CC linker. The peptide may also be linked to a benzylpenicillin molecule  
CC by a glycoamide linker.  
CC Conjugates of the linear peptides and the active agent are particularly  
CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
CC Parkinson's diseases, depression, pain and meningitis, but also for  
CC studying drug behaviour in BBB models.  
XX  
SQ Sequence 18 AA;

Query Match 57.3%; Score 51; DB 21; Length 18;  
Best Local Similarity 62.5%; Pred. No. 0.06;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGELSYSEEFVSVG 17  
|| |||| || ||  
Db 2 ggrlsysrrfststg 17

RESULT 6  
AAY93179



```

XX XX Protegrin peptide sequence.
XX DE
XX KW protegrin; antibiotic; antimicrobial; antiviral; antibacterial;
XX KW antifungal;
XX OS Synthetic.
XX XX
XX XX WO9503325-A.
XX PN
XX PD 02-FEB-1995.
XX XX
XX PF 20-JUL-1994; 94WO-US08305.
XX PR
XX PR 20-JUL-1993; 93US-0093926.
XX PR 26-JUL-1993; 93US-0095769.
XX PR 13-JAN-1994; 94US-0182483.
XX PR 17-MAY-1994; 94US-0243879.
XX XX
XX PA (REGC ) UNIV CALIFORNIA.
XX PI
XX PI Harwig SSL, Kokryakov VN, Lehrer RL;
XX XX
XX DR WPI; 1995-075188/10.
XX XX
XX PT Antibiotic peptide-based cpds. designated protegrin(s) - are
XX PT useful for treating and preventing viral and microbial infections
XX PT and as preservatives
XX XX
XX PS Disclosure; Page 19; 80pp; English.
XX XX
XX CC New peptides are disclosed which are designated "protegrins". The
XX CC peptides are useful as antibacterial, antiviral and antifungal agents in
XX CC both animals and plants. The peptides are 16-18 amino acids in length
XX CC and are characterised by four invariant Cys residues at positions 6, 8,
XX CC 13 and 15 and either (1) by a characteristic pattern of basic and
XX CC hydrophobic amino acids and/or (2) being isolatable from animal (e.g.
XX CC porcine) leukocytes; or analogues of these peptides in which 1-4 of the
XX CC Cys residues is/are replaced by hydrophobic or small amino acids. The
XX CC peptides can be produced synthetically and some can be produced
XX CC recombinantly or can be isolated and purified from their native sources.
XX CC The peptides can be modified by N-acylation and/or C-terminal amidation
XX CC or esterification, and can be in linear or cystine-bridged form. D-Amino
XX CC acid residues can be present.
XX CC The present sequence is a specific example of the protegrin
XX CC analogues in which Cys(8, 15) have been replaced by Ser.
XX XX
XX SQ Sequence 18 AA;

Query Match 51.7%; Score 46; DB 16; Length 18;
Best Local Similarity 56.2%; Pred. No. 0.4;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps

Qy 2 GGELSYSEEFVSVG 17
   |||||::|
Db 2 ggrlcyrkkwcvsrg 17

RESULT 9
AAB30450
ID AAB30450 standard; peptide; 41 AA.
XX XX
XX AC AAB30450;
XX DT
XX DT 06-MAR-2001 (first entry)
XX DE
XX DE Amino acid sequence of a fragment of an Eel galectin family member.
XX XX
XX KW Galectin; Galectin-3; cell growth; tumour; metastasis; cirrhotic liver;
XX KW hepatocellular carcinoma; liver neoplasm.
XX OS Unidentified.
OS OS

```

FT Binding-site /note= "putative substrate binding site"  
 154  
 FT Active-site /note= "putative substrate binding site"  
 155  
 FT Active-site /note= "important for catalytic activity"  
 168..170  
 FT /note= "important for catalytic activity"  
 XX US5683910-A.  
 XX  
 XX 04-NOV-1997.  
 XX  
 XX 13-SEP-1996; 96US-0713828.  
 XX  
 XX 13-SEP-1996; 96US-0713828.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Bandman O, Goli SK;  
 XX WPI; 1997-548985/50.  
 XX N-PSDB; AAT93887.  
 XX New nucleic acid encoding human phosphorylase kinase gamma subunit -  
 PT used for replacement, or gene, therapy of glycogen utilisation  
 PT disorders, also for drug screening and diagnosis  
 XX Claim 1; Fig 1; 25pp; English.  
 XX The present sequence represents a novel human phosphorylase kinase gamma  
 CC subunit, characterised as having homology to other phosphorylase kinase  
 CC gamma subunits. Recombinant human phosphorylase kinase gamma subunit is useful  
 CC for replacement therapy in cases of glycogen utilisation disorders,  
 CC e.g. muscle glycogenosis, causing muscular weakness and atrophy, or  
 CC hypoglycaemia, causing reduced growth and kidney damage. Cells  
 CC transformed with the DNA encoding the present sequence can also be used  
 CC to screen for modulators of human phosphorylase kinase gamma subunit  
 CC activity, i.e. potential therapeutic agents. The protein can also be used  
 CC similarly in gene therapy, while antisense nucleic acid or ribozymes can  
 CC be used to suppress mutant forms of the this gene. The protein can be  
 CC used to raise specific antibodies, useful as immunoassay reagents.  
 CC Fragments of the DNA encoding the present sequence can be used  
 CC diagnostically, e.g. to detect or quantify gene expression, to isolate  
 CC genes or related nucleic acid, to detect activation/induction of human  
 CC phosphorylase kinase gamma subunit-expressing genes and for mapping the  
 CC gene.  
 XX Sequence 307 AA;  
 SQ  
 Query Match 49.4%; Score 44; DB 18; Length 307;  
 Best Local Similarity 90.0%; Pred. No. 24;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 5 LSYSEEEFSV 14  
 |||||  
 Db 189 lsyseeefdv 198  
 RESULT 11  
 AAW76803  
 ID AAW76803 standard; Protein; 307 AA.  
 XX  
 AC AAW76803;  
 XX  
 XX 15-JAN-1999 (first entry)  
 DT  
 XX Human phosphorylase kinase gamma subunit HPHKG protein.  
 DE  
 XX Phosphorylase kinase gamma subunit: HPHKG; human; treatment; glycogen;  
 KW pathological condition; phosphorylase kinase; PHK; hypoglycaemia;  
 KW muscle glycogenosis; atrophy; metabolic abnormality; growth; diagnosis;  
 KW kidney dysfunction; failure; drug screening.

XX Homo sapiens.  
 OS US5833981-A.  
 PN  
 XX  
 XX 10-NOV-1998.  
 PD  
 XX  
 XX 28-AUG-1997; 97US-0919627.  
 PF  
 XX  
 XX 13-SEP-1996; 96US-0713828.  
 PR  
 XX 28-AUG-1997; 97US-0919627.  
 PR  
 XX (INCY-) INCYTE PHARM INC.  
 PA  
 XX Bandman O, Goli SK;  
 PI  
 XX WPI; 1999-008645/01.  
 DR  
 XX N-PSDB; AAV62129.  
 DR  
 XX New isolated human phosphorylase kinase gamma subunit - used to  
 PT develop products for treating e.g. conditions associated with  
 PT diminished glycogen utilisation and consequent glucose depletion  
 XX Claim 1; Fig 1A-B; 25pp; English.  
 PS  
 XX This sequence represents a human phosphorylase kinase gamma subunit,  
 CC HPHKG. The polypeptides and nucleic acids encoding them can be used for  
 CC treating pathological conditions caused by aberrant forms of  
 CC phosphorylase kinase (PHK) and by deficiencies in PHK activity. They can  
 CC be used for treating conditions associated with diminished glycogen  
 CC utilisation and consequent glucose depletion, e.g. muscle glycogenosis,  
 CC involving muscular weakness and atrophy, and hypoglycaemia, which may lead  
 CC to various metabolic abnormalities including poor growth and kidney  
 CC dysfunction or failure. The products can also be used for antibody  
 CC production, detection, diagnosis and drug screening.  
 XX Sequence 307 AA;  
 SQ  
 Query Match 49.4%; Score 44; DB 20; Length 307;  
 Best Local Similarity 90.0%; Pred. No. 24;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 5 LSYSEEEFSV 14  
 |||||  
 Db 189 lsyseeefdv 198  
 RESULT 12  
 AAY52303  
 ID AAY52303 standard; Protein; 307 AA.  
 XX  
 AC AAY52303;  
 XX  
 XX 09-FEB-2000 (first entry)  
 DT  
 XX Novel human phosphorylase kinase (Phk) gamma subunit.  
 DE  
 XX Phosphorylase kinase; gamma subunit; glycogen; calcium dependent;  
 KW catabolism; glucose; mobilisation; glycogen phosphorylase; GP;  
 KW regulation; phosphorylation; activation; genetic heterogeneity; muscle;  
 KW liver; hepatic; glycogenosis; deficiency; disease; exercise intolerance;  
 KW weakness; muscular atrophy; hypoglycaemia; hepatomegaly; antibody;  
 KW detection; treatment.  
 XX Homo sapiens.  
 OS  
 XX US5977320-A.  
 PN  
 XX  
 XX 02-NOV-1999.  
 PD  
 XX  
 XX 11-JUN-1998; 98US-0096245.  
 PF  
 XX

PR 13-SEP-1996; 96US-0713828.  
 PR 28-AUG-1997; 97US-0919627.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Goli SK, Bandman O;  
 PI WPI; 2000-021958/02.  
 DR N-PSDB; AA238239.  
 XX A new antibody to human phosphorylase kinase gamma subunit  
 PT Claim 1; Fig 1; 25pp; English.  
 XX This sequence represents a novel gamma subunit of phosphorylase  
 CC kinase (Phk). Phk regulates glycogen phosphorylase (GP), which is  
 CC responsible for the catabolic breakdown of glycogen to glucose.  
 CC Phk phosphorylates GP, thereby activating it, leading to mobilisation of  
 CC glucose reserves for ATP production. Phk consists of 4 different  
 CC subunits: regulatory alpha and beta subunits, calmodulin (delta subunit)  
 CC and the catalytic gamma subunit. Since the enzyme is composed of 4  
 CC non-identical subunits, Phk deficiency has significant potential for  
 CC genetic heterogeneity. Muscle glycogenosis caused by Phk deficiency  
 CC leads to exercise intolerance, muscle glycogenosis caused by Phk deficiency  
 CC Phk deficiency is also associated with certain hepatic atrophy. Hepatic  
 CC characterised by hypoglycaemia and hepatomegaly. This Phk gamma subunit  
 CC or enzymatically active fragments thereof may be used to raise  
 CC antibodies which may be used to detect, treat and correct  
 CC pathophysiological conditions caused by aberrant forms of Phk, and  
 CC deficiencies in Phk activity. Prior art treatment for Phk deficiency-  
 CC related disease involves dietary regimes which do not prevent the  
 CC course of the disease.  
 XX SQ Sequence 307 AA;

Query Match 49.4%; Score 44; DB 21; Length 307;  
 Best Local Similarity 90.0%; Pred. No. 24;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LSYSEEFVS 14  
 |||||  
 Db 189 lsyseefdv 198

RESULT 13  
 AAY27161  
 ID AAY27161 standard; Protein; 414 AA.

XX AAY27161;

XX 15-SEP-1999 (first entry)

XX Human DRK1 protein.

XX DRK1; DRK2; DAP kinase related apoptosis inducing kinase; human;  
 KW apoptosis; breast cancer; ovarian cancer; lymphoma; autoimmune disease;  
 KW viral infection; adenovirus; poxvirus; HIV; Alzheimer's disease;  
 KW Parkinson's disease; arteriosclerosis; alcoholism; rheumatoid arthritis;  
 KW diabetes.

XX Homo sapiens.

XX WO9933961-A1.

XX 08-JUL-1999.

XX 25-DEC-1998; 98WO-JP05974.

XX 17-APR-1998; 98JP-0108150.

PR 26-DEC-1997; 97JP-0367840.

PR 26-DEC-1997; 97JP-0367841.

PR 17-APR-1998; 98JP-0108149.

XX (ASAH ) ASahi KASEI KOGYO KK.  
 XX Akira S, Kawai T;  
 XX WPI; 1999-430239/36.  
 DR N-PSDB; AAX89196.  
 XX New kinase with apoptosis induction activity useful in the treatment  
 PT of cancer, autoimmune diseases and viral infections  
 XX Claim 2; Page 134-137; 180pp; Japanese.  
 XX The invention provides kinases DRK1 and DRK2 (DAP kinase related  
 CC apoptosis inducing kinase) having apoptosis inducing activity. The  
 CC kinases can be expressed recombinantly by transforming host cells with  
 CC vectors comprising the nucleic acids encoding the kinases. The kinases  
 CC are useful in the treatment, prevention, diagnosis and investigation of  
 CC diseases with which apoptosis is associated, such as hormonally regulated  
 CC cancer (such as breast cancer, ovarian cancer, lymphoma); autoimmune  
 CC diseases; viral infections (such as herpes, adenovirus, poxvirus, HIV);  
 CC Alzheimer's disease; Parkinson's disease; arteriosclerosis; alcoholism;  
 CC rheumatoid arthritis; and diabetes. The present sequence represents the  
 CC human DRK1 amino acid sequence.  
 XX SQ Sequence 414 AA;

Query Match 49.4%; Score 44; DB 20; Length 414;  
 Best Local Similarity 90.0%; Pred. No. 34;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LSYSEEFVS 14  
 |||||  
 Db 280 lsyseefdv 289

RESULT 14  
 AAB65625  
 ID AAB65625 standard; Protein; 414 AA.

XX AAB65625;

XX 27-MAR-2001 (first entry)

XX Novel protein kinase, SEQ ID NO: 151.

XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;  
 KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;  
 KW dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;  
 KW immune disorder; cardiovascular disease; neurodegenerative disease;  
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis.

XX Homo sapiens.

XX WO200073469-A2.

XX 07-DEC-2000.

XX 26-MAY-2000; 2000WO-US14842.

XX 28-MAY-1999; 99US-0136503.

XX (SUGE-) SUGEN INC.

XX Plowman GD, Martinez R, Whyte D, Sudersanam S;

XX WPI; 2001-032161/04.

XX N-PSDB; AAF44651.

XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
 PT treating immune-related diseases and disorders, cardiovascular disease,

PT neurodegenerative diseases and/or cancers -  
PS Claim 10; Fig 1; 310pp; English.  
XX

CC The present sequence is a novel protein kinase. The novel protein kinases  
CC and the nucleic acids that encode them may be used in the treatment and  
CC diagnosis of diseases associated with inappropriate kinase expression  
CC such as immune-related diseases and disorders, cardiovascular disease,  
CC neurodegenerative diseases and/or cancers. The nucleic acids and  
CC complementary sequences may also be used as DNA probes in diagnostic  
CC assays. The kinase polypeptides may be used as antigens in the production  
CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
CC and kinase antagonists may also be used to down regulate kinase  
CC expression and activity. Diseases related to kinase expression and  
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
CC disorders, complications of organ transplantation, myocardial infarction,  
CC immune disorders, cardiomyopathies, strokes, renal failure,  
CC oxidative-stress related disorders, chronic inflammatory bowel disease,  
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,  
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and  
CC reproductive disorders.

XX Sequence 414 AA;

Query Match 49.4%; Score 44; DB 22; Length 414;  
Best Local Similarity 90.0%; Pred. No. 34;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LSYSEEFSSV 14  
|||||||  
Db 280 lsyseeefdv 289

RESULT 15  
AAW36429  
ID AAW36429 standard; peptide; 18 AA.

XX AC AAW36429;

XX DT 13-FEB-1998 (first entry)

XX DE Antimicrobial protegrin peptide (229).

KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
KW retrovirus; HIV; human immunodeficiency virus; preservation;  
KW disinfection; prophylaxis; treatment; infection; disease;  
KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
KW respiratory infection; urinary tract infection; MRSA; protozoan;  
KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
KW methicillin resistant Staphylococcus aureus; systemic candidiasis.

XX OS Synthetic.  
XX Sus scrofa.

XX PN W09718826-A1.

XX PD 29-MAY-1997.

XX PF 22-NOV-1996; 96WO-US18544.

XX PR 21-NOV-1996; 96US-0752852.

XX PR 22-NOV-1995; 95US-0562346.

XX PR 17-MAY-1996; 96US-0649811.

XX PR 01-AUG-1996; 96US-0690921.

XX PA (INTR-) INTRABIOTICS PHARM INC.

XX PA (REGC ) UNIV CALIFORNIA.

XX

PI  
XX

DR WPI; 1997-297871/27.

PT New antimicrobial protegrin peptide(s) - having activity against  
bacteria, yeast, fungi, protozoa and certain strains of viruses  
(e.g. HIV)

XX Claim 23; Page 110; 130pp; English.

XX The present sequence is an antimicrobial protegrin peptide, which  
has a broad spectrum of activity against microbial targets,  
including gram-positive and gram-negative bacteria, yeast, fungi,  
protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
It can be used to preserve or disinfect a variety of materials,  
including medical equipment, foodstuffs, cosmetics, contact lens  
solutions, medicaments or other nutrient containing materials. It  
can also be used for the prophylaxis or treatment of microbial  
infections or diseases in plants and animals, e.g. conjunctivitis,  
keratitis, corneal ulcers, stomach ulcers associated with  
Helicobacter pylori, sexually transmitted diseases, gram-negative  
sepsis, endocarditis, pneumonia and other respiratory infections,  
urinary tract infections, systemic candidiasis and oral mucositis.  
It is biostatic or biocidal against clinically relevant pathogens  
exhibiting multi-drug resistance, e.g. vancomycin resistant  
Enterococcus faecium or faecalis, penicillin resistant  
Streptococcus pneumoniae and methicillin resistant Staphylococcus  
aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
to 1 mg/kg/day, by injection.

XX Sequence 18 AA;

Query Match 48.3%; Score 43; DB 18; Length 18;  
Best Local Similarity 50.0%; Pred. No. 1.3;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGELSYSEEFSSVSG 17

Db 2 ggricyarrfvcvg 17

Search completed: February 12, 2002, 12:30:30  
Job time: 363 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:34 ; Search time 232.64 Seconds  
(without alignments)  
11.317 Million cell updates/sec

Title: US-09-485-571-17  
Perfect score: 89  
Sequence: 1 RGVSVSFRRSYSLRGGR 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues 473505

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_17:  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result NO.	Score	Query Match %	Length	DB	ID	Description
1	46	51.7	192	12	O56033	O56033 human immun
2	45	50.6	192	12	O71966	O71966 human immun
3	45	50.6	192	12	O71971	O71971 human immun
4	45	50.6	448	6	Q9XS63	Q9XS63 equus caball
5	45	50.6	1363	2	Q9F5K8	Q9F5K8 rhodobacter
6	44	49.4	192	12	Q79071	Q79071 human immun
7	44	48.4	192	12	Q9WPB4	Q9WPB4 human immun
8	44	48.4	192	12	Q9E2B2	Q9E2B2 human immun
9	43	48.3	192	12	O11952	O11952 human immun
10	43	48.3	192	12	O11954	O11954 human immun
11	43	48.3	192	12	O55970	O55970 human immun
12	43	48.3	354	2	O30861	O30861 caulobacter
13	42	47.2	141	12	O56039	O56039 human immun
14	42	47.2	185	12	O73433	O73433 human immun
15	42	47.2	192	12	O71266	O71266 human immun
16	42	47.2	192	12	O71276	O71276 human immun
17	42	47.2	192	12	O73432	O73432 human immun
18	42	47.2	192	12	O89450	O89450 human immun
19	42	47.2	192	12	Q80250	Q80250 human immun

20	42	47.2	192	12	Q80251	Q80251 human immun
21	42	47.2	192	12	Q80252	Q80252 human immun
22	42	47.2	192	12	O11955	O11955 human immun
23	42	47.2	192	12	O11964	O11964 human immun
24	42	47.2	192	12	O56016	O56016 human immun
25	42	47.2	192	12	O56017	O56017 human immun
26	42	47.2	192	12	O56018	O56018 human immun
27	42	47.2	192	12	O56019	O56019 human immun
28	42	47.2	192	12	O56020	O56020 human immun
29	42	47.2	192	12	O56023	O56023 human immun
30	42	47.2	192	12	O56030	O56030 human immun
31	42	47.2	192	12	O56031	O56031 human immun
32	42	47.2	192	12	O56034	O56034 human immun
33	42	47.2	192	12	O56038	O56038 human immun
34	42	47.2	192	12	O56040	O56040 human immun
35	42	47.2	192	12	O56041	O56041 human immun
36	42	47.2	192	12	Q9WPC9	Q9WPC9 human immun
37	42	47.2	192	12	Q9WPC8	Q9WPC8 human immun
38	42	47.2	192	12	Q9E2F6	Q9E2F6 human immun
39	42	47.2	192	12	Q9E2F5	Q9E2F5 human immun
40	42	47.2	192	12	Q9E2F4	Q9E2F4 human immun
41	42	47.2	192	12	Q9E2F3	Q9E2F3 human immun
42	42	47.2	192	12	Q9E2F2	Q9E2F2 human immun
43	42	47.2	192	12	Q9E2F1	Q9E2F1 human immun
44	42	47.2	192	12	Q9E2F0	Q9E2F0 human immun
45	42	47.2	192	12	Q9E2E9	Q9E2E9 human immun

## ALIGNMENTS

```

RESULT 1
O56033 PRELIMINARY; PRT; 192 AA.
AC O56033;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE VIF PROTEIN.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses: Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM03.
RX MEDLINE=98105749; PubMed=9445004;
RA Fedavalli V.R., Chappey C., Matala E., Ahmad N.;
RT "Conservation of an intact vif gene of human immunodeficiency virus
   type 1 during maternal-fetal transmission.";
RL J. Virol. 72:1092-1102(1998).
DR EMBL: AF019547; RAC02416.1;
DR InterPro: IPR000475; Viral_infect.
DR Pfam: PF00559; Vif; 1.
DR PRINTS; PRO0349; VIRIONINFECT.
DR ProDom; PD00063; Viral_infect; 1.
SQ SEQUENCE 192 AA; 22454 MW; EB2DC350D41A8A30 CRC64;

```

Query Match 51.7%; Score 46; DB 12; Length 192;  
Best Local Similarity 61.5%; Pred. No. 5.8;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRSYS 13  
| | | | | : | : | | |  
Db 83 RGVSVIEWRRKYS 95

```

RESULT 2
O71966 PRELIMINARY; PRT; 192 AA.
ID O71966;
AC O71966;
DT 01-AUG-1998 (Tremblrel. 07, Created)

```

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VIF PROTEIN.  
 GN VIF.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses: Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HIV-1JC;  
 RA Mwaengo D.M., Novembre F.J.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF049494; AAC68843.1; -  
 DR InterPro: IPR000475; Viral\_infect.  
 DR Pfam: PF00559; Vif; 1.  
 DR PRINTS: PR00349; VIRIONINFECT.  
 DR ProDom: PD000063; Viral\_infect; 1.  
 DR PROSITE: PS00423; GRANINS-2; 1.  
 SQ SEQUENCE 192 AA; 22599 MW; 0D0E1456317A0673 CRC64;

Query Match 50.6%; Score 45; DB 12; Length 192;  
 Best Local Similarity 66.7%; Pred. No. 8.5;  
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GVSVSFRRRSYS 13  
 |||: :|||  
 Db 84 GVSIEWKRKRSYS 95

## RESULT 3

ID O71971 PRELIMINARY; PRT; 192 AA.  
 AC O71971;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VIF PROTEIN.  
 GN VIF.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HIV-1NC;  
 RA Mwaengo D.M., Novembre F.J.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF049495; AAC68852.1; -  
 DR InterPro: IPR000475; Viral\_infect.  
 DR Pfam: PF00559; Vif; 1.  
 DR PRINTS: PR00349; VIRIONINFECT.  
 DR ProDom: PD000063; Viral\_infect; 1.  
 SQ SEQUENCE 192 AA; 22398 MW; A6E790B042ABC996 CRC64;

Query Match 50.6%; Score 45; DB 12; Length 192;  
 Best Local Similarity 66.7%; Pred. No. 8.5;  
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GVSVSFRRRSYS 13  
 |||: :|||  
 Db 84 GVSIEWKRKRSYS 95

## RESULT 4

ID Q9XS63 PRELIMINARY; PRT; 448 AA.  
 AC Q9XS63;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CHROMOGGRANIN A.  
 GN CgA.  
 OS Equus caballus (Horse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ADRENAL MEDULLA;  
 RA Sato F., Ishida N., Hasegawa T., Mukoyama H.;  
 RT "Equine mRNA for chromogranin A";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB025570; BAA76748.1; -  
 DR HSP; P05059; ICfK.  
 DR InterPro: IPR001819; Chromogranin\_AB.  
 DR InterPro: IPR001990; Granin.  
 DR Pfam: PF01271; Granin; 1.  
 DR PRINTS: PR00659; CHROMOGGRANIN.  
 DR PROSITE: PS00422; GRANINS-1; 1.  
 DR PROSITE: PS00423; GRANINS-2; 1.  
 SQ SEQUENCE 448 AA; 49861 MW; EC2D6418F5BA5274 CRC64;

Query Match 50.6%; Score 45; DB 6; Length 448;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RGVSVSFRRSYS 16  
 | : ||| : |  
 Db 361 RSMKLSFRARAYGFRG 376

## RESULT 5

ID Q9F5K8 PRELIMINARY; PRT; 1363 AA.  
 AC Q9F5K8;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE FLAGELLAR HOOK-ASSOCIATED PROTEIN 1.  
 GN FLGK.  
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 OX NCBI\_TaxID=1063;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gonzalez-Pedrajo B., De La Mora J., Ballado T., Camarena L.,  
 RA Dreyfus G.;  
 RT "Isolation and Complementatation of a Flagellar P-ring Mutant of  
 RT Rhodobacter sphaeroides";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF317649; AAG31286.1; -  
 DR InterPro: IPR001444; Flag\_bb\_rod.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR Pfam: PF00460; flg\_bb\_rod; 1.  
 DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SQ SEQUENCE 1363 AA; 133154 MW; AD7BB4856117808C CRC64;

Query Match 50.6%; Score 45; DB 2; Length 1363;  
 Best Local Similarity 57.9%; Pred. No. 67;  
 Matches 11; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 RGVSVSFRRSYS 17  
 | ||| : |||  
 Db 973 RSVSVSYEGQSYTLRWTGG 991

## RESULT 6

ID Q79071 PRELIMINARY; PRT; 192 AA.  
 AC Q79071;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE VIF (VIRAL INFECTIVITY FACTOR).  
GN Human immunodeficiency virus type 1.  
OS Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OC NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT A10;  
RX MEDLINE=94303241; PubMed=8030283;  
RA Wieland U., Hartmann J., Suhr H., Salzberger B., Eggers H.J.,  
RA Kuehn J.E.;  
RT "In vivo genetic variability of the HIV-1 vif gene.";  
RL Virology 203:43-51(1994).  
DR EMBL: Z30679; CAA83159.1; -;  
DR InterPro: IPR000475; Viral\_infect.  
DR Pfam: PF00559; Vif; 1.  
DR PRINTS: PR00349; VIRIONINFECT.  
DR ProDom: PD000063; Viral\_infect; 1.  
SQ SEQUENCE 192 AA; 22504 MW; 3163F6B8524974F7 CRC64;

Query Match 49.4%; Score 44; DB 12; Length 192;  
Best Local Similarity 47.1%; Pred. No. 12;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYSLRG 17  
:||||:|:|:|:|  
Db 83 QGVSIEMRRKRYSTQVG 99

RESULT 7  
Q9WPB4 PRELIMINARY; PRT; 192 AA.  
ID Q9WPB4;  
AC Q9WPB4;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE VIF PROTEIN.  
GN VIF.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VI05002B1H;  
RA Hassaine G., Agostini I., Candotti D., Bessou G., Caballero M.,  
RA Agut H., Autran B., Barthalay Y., The French ALT Study Group,  
RA Vigne R.;  
RT "Characterization of human immunodeficiency virus type 1 vif gene in  
RT long-term asymptomatic individuals.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF143128; AAD37895.1; -;  
DR InterPro: IPR000475; Viral\_infect.  
DR Pfam: PF00559; Vif; 1.  
DR PRINTS: PR00349; VIRIONINFECT.  
DR ProDom: PD000063; Viral\_infect; 1.  
SQ SEQUENCE 192 AA; 22506 MW; 5A876245CFA996C9 CRC64;

Query Match 49.4%; Score 44; DB 12; Length 192;  
Best Local Similarity 47.1%; Pred. No. 12;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYSLRG 17  
:||||:|:|:|:|  
Db 83 QGVSIEMRRKRYSTQVG 99

RESULT 8  
Q9E2B2 PRELIMINARY; PRT; 192 AA.  
ID Q9E2B2;  
AC Q9E2B2;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE VIF PROTEIN (FRAGMENT).  
GN VIF.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WNT3.5F.16;  
RA Ahmad N., Yedavalli V.S.R.K.;  
RT "Low Conservation of Functional Domains of Human Immunodeficiency  
RT Virus Type 1 vif and vpr Genes Correlates with Lack of Vertical  
RT Transmission.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF275040; AAG32213.1; -;  
DR InterPro: IPR000475; Viral\_infect.  
DR Pfam: PF00559; Vif; 1.  
DR PRINTS: PR00349; VIRIONINFECT.  
DR ProDom: PD000063; Viral\_infect; 1.  
FT NON\_TER 192  
SQ SEQUENCE 192 AA; 22726 MW; BA38FB592232DA07 CRC64;

Query Match 49.4%; Score 44; DB 12; Length 192;  
Best Local Similarity 47.1%; Pred. No. 12;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYSLRG 17  
:||||:|:|:|:|  
Db 83 QGVSIEMRRKRYSTQVG 99

RESULT 9  
O11952 PRELIMINARY; PRT; 192 AA.  
ID O11952;  
AC O11952;  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE VIF 23 KDA PROTEIN.  
GN VIF.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97068252; PubMed=8911580;  
RA Tomimaga K., Kato S., Negishi M., Takano T.;  
RT "A high frequency of defective vif genes in peripheral blood  
RT mononuclear cells from HIV type 1-infected individuals.";  
RL AIDS Res. Hum. Retroviruses 12:1543-1549(1996).  
DR EMBL: D70861; BAA20303.1; -;  
DR InterPro: IPR000475; Viral\_infect.  
DR Pfam: PF00559; Vif; 1.  
DR PRINTS: PR00349; VIRIONINFECT.  
DR ProDom: PD000063; Viral\_infect; 1.  
SQ SEQUENCE 192 AA; 22595 MW; 5B4CF50CCB63E9BE CRC64;

Query Match 48.3%; Score 43; DB 12; Length 192;  
Best Local Similarity 53.8%; Pred. No. 18;  
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYSLRG 13  
:||||:|:|:|:|  
Db 83 QGVSIEMRRKRYSTQVG 95

RESULT 10  
O11954 PRELIMINARY; PRT; 192 AA.  
ID O11954;  
AC O11954;

```

DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE VIF 23 KDA PROTEIN.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97069252; PubMed=8911580;
RA Tomimaga K., Kato S., Negishi M., Takano T.;
RT "A high frequency of defective vif genes in peripheral blood
RT mononuclear cells from HIV type 1-infected individuals.";
RL AIDS Res. Hum. Retroviruses 13:1543-1549(1996).
DR EMBL; D70864; BAA20305.1; -
DR InterPro; IPR000475; Viral_infect.
DR Pfam; PF00559; Vif; 1.
DR PRINTS; PR00349; VIRIONINFECT.
DR ProDom; PD000063; Viral_infect; 1.
SQ SEQUENCE 192 AA; 22578 MW; 30251A6F286C165E CRC64;

Query Match 48.3%; Score 43; DB 12; Length 192;
Best Local Similarity 53.8%; Pred. No. 18;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRSYS 13
||||:|:|:|
DB 83 RGVSIWKRKRGYN 95

RESULT 11
O55970
ID O55970 PRELIMINARY; PRT; 192 AA.
AC O55970;
DT -01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE VIF PROTEIN.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BM07;
RX MEDLINE=98105749; PubMed=9445004;
RA Yedavalli V.R., Chappey C., Matala E., Ahmad N.;
RT "Conservation of an intact vif gene of human immunodeficiency virus
RT type 1 during maternal-fetal transmission.";
RL J. Virol. 72:1092-1102(1998).
DR EMBL; AF019469; AAC02348.1; -
DR InterPro; IPR000475; Viral_infect.
DR Pfam; PF00559; Vif; 1.
DR PRINTS; PR00349; VIRIONINFECT.
DR ProDom; PD000063; Viral_infect; 1.
SQ SEQUENCE 192 AA; 22507 MW; D9A8F709DC2912D3 CRC64;

Query Match 48.3%; Score 43; DB 12; Length 192;
Best Local Similarity 53.8%; Pred. No. 18;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGVSVSFRRSYS 13
||||:|:|:|
DB 83 RGVSIWKRKRTYS 95

RESULT 12
O30861
ID O30861 PRELIMINARY; PRT; 354 AA.
AC O30861;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GTP-BINDING PROTEIN (GTP-BINDING PROTEIN CGTA).
GN CGTA OR CC0315.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NA1000;
RX MEDLINE=97474268; PubMed=9335292;
RA Maddock J., Bhatt A., Koch M., Skidmore J.;
RT "Identification of an essential Caulobacter crescentus gene encoding a
RT member of the Opg family of GTP-binding proteins.";
RL J. Bacteriol. 179:6426-6431(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AF019407; AAB81507.1; -
DR EMBL; AE005705; AAK22302.1; -
DR TIGR; CC0315; -
DR InterPro; IPR000765; GTP1_OBG.
DR Pfam; PF01018; GTP1_OBG; 1.
DR PRINTS; PR00326; GTP1_OBG.
DR PROSITE; PS00905; GTP1_OBG; 1.
KW Complete proteome.
SQ SEQUENCE 354 AA; 37871 MW; DF81EAC1012A5266 CRC64;

Query Match 48.3%; Score 43; DB 2; Length 354;
Best Local Similarity 62.5%; Pred. No. 35;
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GVSVSFRRSYSLRGG 17
| | | | | | | | | |
DB 18 GGSVSFRREKYEYGG 33

RESULT 13
O56039
ID O56039 PRELIMINARY; PRT; 141 AA.
AC O56039;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE VIF PROTEIN (FRAGMENT).
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CM09;
RX MEDLINE=98105749; PubMed=9445004;
RA Yedavalli V.R., Chappey C., Matala E., Ahmad N.;
RT "Conservation of an intact vif gene of human immunodeficiency virus
RT type 1 during maternal-fetal transmission.";
RL J. Virol. 72:1092-1102(1998).
DR EMBL; AF019553; AAC02422.1; -
DR InterPro; IPR000475; Viral_infect.
DR Pfam; PF00559; Vif; 1.
DR PRINTS; PR00349; VIRIONINFECT.

```

DR ProDom; PD000063; Viral\_infect; 1.  
FT NON\_TER 141  
SQ SEQUENCE 141 AA; 16756 MW; 38D693B7A32F094F CRC64;

Query Match 47.2%; Score 42; DB 12; Length 141;  
Best Local Similarity 53.8%; Pred. No. 19;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|||: :|:| ||  
Db 83 QGVSIWRRRKYS 95

## RESULT 14

Q73433 PRELIMINARY; PRT; 185 AA.  
AC Q73433;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE VIRUS INFECTIVITY FACTOR.  
GN VIF.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95191036; PubMed=7884906;  
RA Sovia P., van Ranst M., Gupta P., Balachandran R., Chao W., Itescu S.,  
RA McKinley G., Volsky D.J.;  
RT "Conservation of an intact human immunodeficiency virus type 1 vif  
gene in vitro and in vivo.";  
RL J. Virol. 69:2557-2564(1995).  
DR EMBL: U42253; AAA83802.1; -;  
DR InterPro: IPR000475; Viral\_infect.  
DR Pfam: PF00559; Vif; 1.  
DR ProDom; PD000063; Viral\_infect; 1.  
SQ SEQUENCE 185 AA; 21449 MW; D73CFA994D541E2A CRC64;

Query Match 47.2%; Score 42; DB 12; Length 185;  
Best Local Similarity 53.8%; Pred. No. 26;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|||: :|:| ||  
Db 76 QGVSIWRRRKYS 88

## RESULT 15

O71266 PRELIMINARY; PRT; 192 AA.  
AC O71266;  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE VIF PROTEIN.  
GN VIF.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MBCC54;  
RA Oelrichs R.B., McPhee D.A., Deacon N.J.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF042103; AAD03218.1; -;  
DR InterPro: IPR000475; Viral\_infect.  
DR Pfam: PF00559; Vif; 1.  
DR PRINTS: PR00349; VIRIONINFECT.  
DR ProDom; PD000063; Viral\_infect; 1.  
SQ SEQUENCE 192 AA; 22511 MW; E052974F0BC1DA27 CRC64;

Query Match 47.2%; Score 42; DB 12; Length 192;  
Best Local Similarity 53.8%; Pred. No. 27;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|||: :|:| ||  
Db 83 QGVSIWRRRKYS 95

Search completed: February 12, 2002, 12:38:35  
Job time: 748 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:48 ; Search time 57.2 seconds  
(without alignments)  
9.821 Million cell updates/sec

Title: US-09-485-571-17

Perfect score: 89

Sequence: 1 RGVSVFRRRSYSLRGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	43	48.3	449	1	CMGA_BOVIN
2	41	46.1	188	1	VIF_HV126
3	41	46.1	192	1	VIF_HV1B5
4	41	46.1	192	1	VIF_HV1EL
5	41	46.1	192	1	VIF_HV1ND
6	41	46.1	192	1	VIF_HV1RH
7	41	46.1	192	1	VIF_HV1Z2
8	41	46.1	202	1	RL13_T0BAC
9	41	46.1	360	1	Y303_HELPJ
10	41	46.1	360	1	Y303_HELPY
11	40	44.9	334	1	Y389_BUCAI
12	40	44.9	383	1	INSI_ECOLI
13	40	44.9	390	1	YHBZ_ECOLI
14	40	44.9	446	1	PFES_PSEAE
15	40	44.9	457	1	CMGA_HUMAN
16	40	44.9	509	1	VE2_HPV36
17	40	44.9	854	1	VG12_BP2A
18	40	44.9	854	1	VG12_BP2A
19	39	43.8	60	1	HSP1_ORNAN
20	39	43.8	184	1	RL13_SCHMA
21	39	43.8	423	1	P65_MYCPN
22	39	43.8	452	1	PUR9_THEMEA
23	39	43.8	463	1	CMGA_MOUSE
24	39	43.8	466	1	CMGA_RAT
25	39	43.8	520	1	C6V1_DROME
26	39	43.8	860	1	VG12_BPB03
27	39	43.8	866	1	FIBA_HUMAN
28	39	43.8	1247	1	NOS_ANOST
29	38	42.7	60	1	HSP1_DASVI
30	38	42.7	61	1	HSP1_ANTLA
31	38	42.7	61	1	HSP1_ANTSW
32	38	42.7	61	1	HSP1_PARBI
33	38	42.7	61	1	HSP1_SARHA

#### RESULT 1

ID	CMGA_BOVIN	STANDARD;	PRT;	449 AA.
AC	P05059;	P79392;		
DT	13-AUG-1987	(Rel. 05, Created)		
DT	01-NOV-1988	(Rel. 09, Last sequence update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	CHROMOGHRANIN A PRECURSOR (CGA) (PITUITARY SECRETORY PROTEIN I) (SP-I)			
DE	[CONTAINS: VASOSTATIN-1; CHROMOSTATIN; CHROMACIN; PANCREASTATIN; WE-			
DE	14; CATESTATIN].			
GN	CHGA.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92140395; PubMed=1779968;			
RA	Iacangelo A.L., Grimes M., Eiden L.E.;			
RT	"The bovine chromogranin A gene: structural basis for hormone			
RT	regulation and generation of biologically active peptides.";			
RL	Mol. Endocrinol. 5:1651-1660(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86300648; PubMed=3755681;			
RA	Benedict U.M., Baeuerle P.A., Konecki D.S., Frank R., Powell J.,			
RA	Mallet J., Huttner W.B.;			
RT	"The primary structure of bovine chromogranin A: a representative of			
RT	a class of acidic secretory proteins common to a variety of			
RT	peptidergic cells.";			
RL	EMBO J. 5:1495-1502(1986).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86311345; PubMed=3018587;			
RA	Iacangelo A., Affolter H.-U., Eiden L.E., Herbert E., Grimes M.;			
RT	"Bovine chromogranin A sequence and distribution of its messenger RNA			
RT	in endocrine tissues.";			
RL	Nature 323:82-86(1986).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87260925; PubMed=3474638;			
RA	Ahn T.G., Cohn D.V., Gorr S.O., Ornstein D.L., Kashdan M.A.,			
RA	Levine M.A.;			
RT	"Primary structure of bovine pituitary secretory protein I			
RT	(chromogranin A) deduced from the cDNA sequence.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 84:5043-5047(1987).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97228583; PubMed=9074643;			
RA	Kang Y.K., Yoo S.H.;			
RT	"Identification of the secretory vesicle membrane binding region of			
RT	chromogranin A.";			
RL	FEBS Lett. 404:87-90(1997).			
RN	[6]			
RP	SEQUENCE OF 19-45, AND CALCIUM-BINDING.			

P42134 dasykaluta  
P42140 murexia lon  
P42129 antechinus  
Q70623 human immun  
P03401 human immun  
P05898 human immun  
P12504 human immun  
P31820 human immun  
P35964 human immun  
P41128 brassica na  
P41129 brassica na  
P41127 arabidopsis

#### ALIGNMENTS

RX MEDLINE-90354431; PubMed-2387861;  
RA Yoo S.H., Albanesi J.P.;  
RT "Ca2(+)-induced conformational change and aggregation of chromogranin  
A.";  
RL J. Biol. Chem. 265:14414-14421(1990).  
RN [7]  
RP SEQUENCE OF 142-161, AND SYNTHESIS OF CHROMOSTATIN.  
RX MEDLINE-91142185; PubMed-1996343;  
RA Galindo E., Rill A., Bader M.-F., Aunis D.;  
RT "Chromostatin, a 20-amino acid peptide derived from chromogranin A,  
inhibits chromaffin cell secretion.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:1426-1430(1991).  
RN [8]  
RP ERRATUM.  
RA Galindo E., Rill A., Bader M.-F., Aunis D.;  
RL Proc. Natl. Acad. Sci. U.S.A. 91:832-832(1994).  
RN [9]  
RP SEQUENCE OF 266-312.  
RX MEDLINE-89331945; PubMed-2756155;  
RA Nakano I., Funakoshi A., Miyasaka K., Ishida K., Makk G., Angwin P.,  
RT "Isolation and characterization of bovine pancreastatin.";  
RL Regul. Pept. 25:207-213(1989).  
RN [10]  
RP SEQUENCE OF 191-212 (CHROMACIN).  
RC TISSUE-Chromaffin granules;  
RX MEDLINE-97067080; PubMed-8910482;  
RA Strub J.-M., Goumon Y., Lugardon K., Capon C., Lopez M., Moniatte M.,  
RT "Antibacterial activity of glycosylated and phosphorylated  
chromogranin A-derived peptide 173-194 from bovine adrenal medullary  
chromaffin granules";  
RL J. Biol. Chem. 271:28533-28540(1996).  
RN [11]  
RP CHARACTERIZATION OF CATESTATIN.  
RX MEDLINE-97439785; PubMed-9294131;  
RA Mahata S.K., O'Connor D.T., Mahata M., Yoo S.H., Taupenot L., Wu H.,  
RA Gill B.M., Farmer R.J.;  
RT "Novel autocrine feedback control of catecholamine release. A discrete  
chromogranin A fragment is a noncompetitive nicotinic cholinergic  
antagonist.";  
RL J. Clin. Invest. 100:1623-1633(1997).  
RN [12]  
RP CHARACTERIZATION OF CATESTATIN.  
RX MEDLINE-99000113; PubMed-9786174;  
RA Kennedy B.P., Mahata S.K., O'Connor D.T., Ziegler M.G.;  
RT "Mechanism of cardiovascular actions of the chromogranin A fragment  
catestatin in vivo.";  
RL Peptides 19:1241-1248(1998).  
RN [13]  
RP 3D-STRUCTURE MODELING OF CATESTATIN.  
RX MEDLINE-99025667; PubMed-9809795;  
RA Tsigelny I., Mahata S.K., Taupenot L., Preece N.E., Mahata M.,  
RA Khan I., Farmer R.J., O'Connor D.T.;  
RT "Mechanism of action of chromogranin A on catecholamine release:  
molecular modeling of the catestatin region reveals a beta-  
strand/loop/beta-strand structure secured by hydrophobic interactions  
and predictive of activity";  
RL Regul. Pept. 77:43-53(1998).  
RN [14]  
RP CHARACTERIZATION OF VASOSTATIN-1.  
RX MEDLINE-20219105; PubMed-10753865;  
RA Lugardon K., Raffner R., Goumon Y., Corti A., Delmas A., Bulet P.,  
RA Aunis D., Metz-Boutigue M.-H.;  
RT "Antibacterial and antifungal activities of vasostatin-1, the N-  
terminal fragment of chromogranin A.";  
RL J. Biol. Chem. 275:10745-10753(2000).  
RN [15]  
RP CARBOHYDRATE-BINDING SITES, PHOSPHORYLATION, AND DISULFIDE BOND.  
RX MEDLINE-99459228; PubMed-10527498;  
RA Bauer S.H., Zhang X.Y., Van Dongen W., Claeys M., Przybylski M.;  
RT "Chromogranin A from bovine adrenal medulla: molecular  
characterization of glycosylations, phosphorylations, and sequence

heterogeneities by mass spectrometry.";  
Anal. Biochem. 274:69-80(1999).  
RT FUNCTION: PANCREASTATIN STRONGLY INHIBITS GLUCOSE INDUCED INSULIN  
RELEASE FROM THE PANCREAS.  
CC FUNCTION: CHROMOSTATIN COMPLETELY INHIBITS CATECHOLAMINE RELEASE  
FROM CHROMAFFIN CELLS.  
CC FUNCTION: CHROMACIN HAS ANTIBACTERIAL ACTIVITY AGAINST MICROCOCCUS  
LUTEUS.  
CC FUNCTION: CATESTATIN INHIBITS CATECHOLAMINE RELEASE FROM  
CHROMAFFIN CELLS AND NORADRENERGIC NEURONS BY ACTING AS A NON-  
COMPETITIVE NICOTINIC CHOLINERGIC ANTAGONIST.  
CC FUNCTION: VASOSTATIN-1 HAS ANTIBACTERIAL ACTIVITY AGAINST GRAM-  
POSITIVE BACTERIA M.LUTEUS, B. MEGATERIUM, NOT ACTIVE AGAINST GRAM-  
S.AUREUS AND L.MONOCOCCOGENES AND AGAINST GRAM-NEGATIVE BACTERIA  
E.COLI, E.CLOACAE, S.TYPHIMURIUM, K.PNEUMONIAE AND P.AURUGINOSA.  
CC POSSESSES ANTIFUNGAL ACTIVITY AGAINST N.CRASSA, A.FUMIGATUS,  
A.BRASSICOLA, N.HEMATOCOCCA, F.CULMORUM AND F.OXYSPORUM AND  
AGAINST S.CEREVIASE AND C.ALBCANS YEAST. INACTIVE AGAINST  
T.MENTAGROPHYTES.  
CC SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY  
GRANULES.  
CC MASS SPECTROMETRY: MW=8584.9; METHOD=MALDI; RANGE=19-94.  
CC MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
CC SIMILARITY: BELONGS TO THE CHROMOGANIN / SECRETOGANIN PROTEIN  
FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; S79270; AAB21297.1; JOINED.  
EMBL; S79256; AAB21297.1; JOINED.  
EMBL; S79258; AAB21297.1; JOINED.  
EMBL; S79260; AAB21297.1; JOINED.  
EMBL; S79262; AAB21297.1; JOINED.  
EMBL; S79264; AAB21297.1; JOINED.  
EMBL; S79266; AAB21297.1; JOINED.  
EMBL; S79268; AAB21297.1; JOINED.  
EMBL; X04012; CAA27636.1; JOINED.  
EMBL; X04298; CAA27841.1; JOINED.  
EMBL; M16971; AAA30765.1; JOINED.  
EMBL; U73523; AAC48700.1; JOINED.  
PIR; A41520; A41520.  
PIR; A24175; A24175.  
PIR; A28033; A28033.  
PDB; 1CFK; 22-MAR-99.  
GlycoSuiteDB; P05059; CHROMOGANIN\_AB.  
InterPro; IPR001819; Chromogranin\_AB.  
InterPro; IPR001990; Granin.  
Pfam; PF01271; Granin; 1.  
PRINTS; PR00659; CHROMOGANIN.  
PROSITE; PS00422; GRANINS\_1; 1.  
PROSITE; PS00423; GRANINS\_2; 1.  
Signal; Amidation; Glycoprotein; Calcium-binding; Phosphorylation;  
Polymorphism; 3D-structure; Antibiotic; Fungicide.  
SIGNAL 1 18  
CHAIN 19 449 CHROMOGANIN A.  
PEPTIDE 19 94 VASOSTATIN-1.  
PEPTIDE 142 161 CHROMOSTATIN.  
PEPTIDE 191 212 CHROMACIN.  
PEPTIDE 266 312 PANCREASTATIN.  
PEPTIDE 334 347 WE-14.  
PEPTIDE 362 382 CATESTATIN.  
MOD\_RES 99 99 PHOSPHORYLATION (PARTIAL).  
MOD\_RES 142 142 PHOSPHORYLATION (PARTIAL).  
MOD\_RES 191 191 PHOSPHORYLATION (PARTIAL).  
MOD\_RES 315 315 PHOSPHORYLATION (PARTIAL).  
MOD\_RES 390 390 PHOSPHORYLATION.

```
Query Match      48.3%; Score 43; DB 1; Length 449;
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYSLRG 16
DB 362 RSMRLSFRARGYGFRG 377

RESULT 2
VIF_HV126
ID VIF_HV126 STANDARD; PRT; 188 AA.
AC P04596;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTION FACTOR (SOR PROTEIN).
GN VIF.
OS Human immunodeficiency virus type 1 (Zaire 6 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87248097; PubMed=3036660;
RA Srinivasan A., Anand R., York D., Ranganathan P., Feorino P.,
RA Schochetman G., Curran J., Kalyanaram V.S., Luciw P.A.,
RA Sanchez-Pescador R.;
RT "Molecular characterization of human immunodeficiency virus from
RT Zaire: nucleotide sequence analysis identifies conserved and variable
RT domains in the envelope gene.";
RL Gene 52:71-82(1987).
CC -!- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K03458; AAA45379.1; -
CC PIR; B26192; ASLJZR.
CC HIV; K03458; VIF$26.
CC InterPro: IPR000475; Viral_infect.
CC Pfam: PF00559; Vif; 1.
CC PRINTS; PR00349; VIRIONINFECT.
CC ProDom; PD000063; Viral_infect; 1.
CC AIDS.
CC SEQUENCE 188 AA; 22159 MW; E3B67D89E0ACDB4F CRC64;

Query Match      46.1%; Score 41; DB 1; Length 188;
Best Local Similarity 53.8%; Pred. No. 3.3;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13
DB 83 QGVSIWRRRRYS 95

RESULT 3
VIF_HV1B5
ID VIF_HV1B5 STANDARD; PRT; 192 AA.
AC P04598;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTION FACTOR (SOR PROTEIN).
GN VIF.
OS Human immunodeficiency virus type 1 (BH5 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
```

```
OX NCBI_TaxID=11682;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85111123; PubMed=2578615;
RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R.,
RA Josephs S.F., Doran E.R., Rafalski J.A., Whitehorn E.A.,
RA Baumeister K., Ivanoff L., Petteway S.R. Jr., Pearson M.L.,
RA Lautenberger J.A., Papas T.S., Ghayab J., Chang N.T., Gallo R.C.,
RA Wong-Staal F.;
RT "Complete nucleotide sequence of the AIDS virus, HTLV-III.";
RL Nature 313:277-284(1985).
CC -!- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02012; AAA44654.1; -
CC HIV; K02012; VIF$BH5.
CC InterPro: IPR000475; Viral_infect.
CC Pfam: PF00559; Vif; 1.
CC PRINTS; PR00349; VIRIONINFECT.
CC ProDom; PD000063; Viral_infect; 1.
CC AIDS.
CC SEQUENCE 192 AA; 22520 MW; AC17E169F5354493 CRC64;

Query Match      46.1%; Score 41; DB 1; Length 192;
Best Local Similarity 53.8%; Pred. No. 3.4;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13
DB 83 QGVSIWRRRRYS 95

RESULT 4
VIF_HV1EL
ID VIF_HV1EL STANDARD; PRT; 192 AA.
AC P04597;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTION FACTOR (SOR PROTEIN).
GN VIF.
OS Human immunodeficiency virus type 1 (ELI isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86245056; PubMed=2424612;
RA Alizon M., Wain-Hobson S., Montagnier L., Sonigo P.;
RT "Genetic variability of the AIDS virus: nucleotide sequence analysis
RT of two isolates from African patients.";
RL Cell 46:63-74(1986).
CC -!- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K03454; AAA44326.1; -
CC EMBL; A07108; CAA00613.1; -
CC HIV; K03454; VIF$ELI.
CC InterPro: IPR000475; Viral_infect.
```

DR Pfam; PF00559; Vif; 1.  
 DR PRINTS; PR00349; VIRIONINFECT.  
 DR ProDom; PD000063; Viral\_infect; 1.  
 KW AIDS.  
 SQ SEQUENCE 192 AA; 22689 MW; 169395846CCA2082 CRC64;

Query Match 46.1%; Score 41; DB 1; Length 192;  
 Best Local Similarity 53.8%; Pred. NO. 3.4;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
 :|||: :||| :||  
 Db 83 QGVSIWKRKRRYS 95

## RESULT 5

ID VIF\_HVIND STANDARD; PRT; 192 AA.  
 AC P18805;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 DE VIRION INFECTION FACTOR (SOR PROTEIN).  
 GN VIF.

OS Human immunodeficiency virus type 1 (NDK isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11695;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=90034200; PubMed=2806917;  
 RA Spire B., Sire J., Zachar V., Rey F., Barre-Sinoussi F., Galibert F.,  
 RA Hampe A., Chermann J.C.;  
 RT "Nucleotide sequence of HIV-1 NDK: a highly cytopathic strain of the  
 RT human immunodeficiency virus.";  
 RL Gene 81:275-284 (1989).

CC -!- FUNCTION: DETERMINES VIRUS INFECTION.

CC -!- MISCELLANEOUS: NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH  
 CC AIDS, AND IS A HIGHLY CYTOPATHIC STRAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M27323; AAA44870.1; -  
 CC PIR; JQ0069; ASLJND.  
 CC HIV; M27323; VIF\$NDK.

DR InterPro; IPR000475; Viral\_infect.  
 DR Pfam; PF00559; Vif; 1.  
 DR PRINTS; PR00349; VIRIONINFECT.  
 DR ProDom; PD000063; Viral\_infect; 1.  
 KW AIDS.  
 SQ SEQUENCE 192 AA; 22556 MW; 15B9BDC424496D22 CRC64;

Query Match 46.1%; Score 41; DB 1; Length 192;  
 Best Local Similarity 53.8%; Pred. NO. 3.4;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
 :|||: :||| :||  
 Db 83 QGVSIWKRKRRYS 95

## RESULT 6

VIF\_HV1RH  
 ID VIF\_HV1RH STANDARD; PRT; 192 AA.  
 AC P05900;  
 DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 DE VIRION INFECTION FACTOR (SOR PROTEIN).  
 GN VIF.  
 OS Human immunodeficiency virus type 1 (RF/HAT isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11701;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Starck B.R., Hahn B.H., Shaw G.M., McNeely P.D., Modrow S.,  
 RA Wolf H., Parks E.S., Parks W.P., Josephs S.F., Gallo R.C.,  
 RA Wong-Staal F.;  
 RL Submitted (XXX-1987) to the HIV data bank.  
 CC -!- FUNCTION: DETERMINES VIRUS INFECTION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M17451; AAA45054.1; -  
 CC HIV; M17451; VIF\$RF.  
 CC InterPro; IPR000475; Viral\_infect.  
 CC Pfam; PF00559; Vif; 1.  
 CC PRINTS; PR00349; VIRIONINFECT.  
 CC ProDom; PD000063; Viral\_infect; 1.  
 KW AIDS.  
 SQ SEQUENCE 192 AA; 22687 MW; F005E0AE621A5C6C CRC64;

Query Match 46.1%; Score 41; DB 1; Length 192;  
 Best Local Similarity 53.8%; Pred. NO. 3.4;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
 :|||: :||| :||  
 Db 83 QGVSIWKRKRRYS 95

## RESULT 7

VIF\_HV1Z2  
 ID VIF\_HV1Z2 STANDARD; PRT; 192 AA.  
 AC P12503;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 DE VIRION INFECTION FACTOR (SOR PROTEIN).  
 GN VIF.

OS Human immunodeficiency virus type 1 (Z2/CDC-234 isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11683;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Theodore T., Buckler-White A.;  
 RL Submitted (NOV-1988) to the HIV data bank.  
 CC -!- FUNCTION: DETERMINES VIRUS INFECTION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M22639; AAA45367.1; -  
 CC HIV; M22639; VIF\$Z226.  
 CC InterPro; IPR000475; Viral\_infect.  
 CC Pfam; PF00559; Vif; 1.  
 CC PRINTS; PR00349; VIRIONINFECT.

```
DR ProDom: PD000063; Viral_infect; 1.
KW AIDS.
SQ SEQUENCE 192 AA; 22572 MW; 0294A76114C7C643 CRC64;

Query Match 46.1%; Score 41; DB 1; Length 192;
Best Local Similarity 53.8%; Pred. No. 3.4;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYS 13
   :|||: :|:| |
Db 83 QGVSVFRRRSYS 95

RESULT 8
ID RL13_TOBAC STANDARD; PRT; 202 AA.
AC P49627;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L13 (CLONE 6.2.1).
GN RPL13
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. XANTHI; TISSUE=Seedling;
RX MEDLINE=95023159; PubMed=7937121;
RA Estruch J.J., Crossland L., Goff S.A.;
RT "Plant activating sequences: positively charged peptides are
RT functional as transcriptional activation domains.";
RL Nucleic Acids Res. 22:3983-3989(1994).
CC -|- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L31416; AAA72054.1;
DR InterPro: IPR001380; Ribosomal_L13E.
DR Pfam: PF01294; Ribosomal_L13e; 1.
DR ProDom: PD004443; Ribosomal_L13E; 1.
DR PROSITE: PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
SQ SEQUENCE 202 AA; 23391 MW; 522CB43CCD80A67E CRC64;

Query Match 46.1%; Score 41; DB 1; Length 202;
Best Local Similarity 53.3%; Pred. No. 3.6;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 GVSVSFRRRSYSLRG 16
   :|:| :|:| :|:|
Db 89 GIAVDRRRRRNRSLG 103

RESULT 9
Y303_HELPJ STANDARD; PRT; 360 AA.
ID Y303_HELPJ
AC Q9ZMD3;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE GTP-BINDING PROTEIN JHP0288.
GN JHP0288.

OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120537; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -|- SIMILARITY: BELONGS TO THE GTP1 / OBG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001465; AAD05858.1;
DR InterPro: IPR000765; GTP1_OBG.
DR Pfam: PF01018; GTP1_OBG; 1.
DR PRINTS: PR00326; GTP1_OBG.
DR PROSITE: PS00905; GTP1_OBG; 1.
KW Hypothetical protein; GTP-binding; Complete proteome.
FT NP_BIND 163 170 GTP (BY SIMILARITY).
FT NP_BIND 210 214 GTP (BY SIMILARITY).
FT NP_BIND 279 282 GTP (BY SIMILARITY).
SQ SEQUENCE 360 AA; 38872 MW; 4C558B6479AA38FF CRC64;

Query Match 46.1%; Score 41; DB 1; Length 360;
Best Local Similarity 53.8%; Pred. No. 6.7;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VSFRRRSYSLRGG 17
   :|:|:| :|:|
Db 20 VSFRRRKFVIRGG 32

RESULT 10
Y303_HELPY STANDARD; PRT; 360 AA.
ID Y303_HELPY
AC Q25074;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE GTP-BINDING PROTEIN HP0303.
GN HP0303.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Klenk H.-P., Gill S., Dougherty B.A.,
RA Fleischmann R.D., Ketchum K.A., Zhou L., Kirkness E.F., Peterson S.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Glodek A.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Hickey E.K.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey J.M.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathley L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
```

```

RT pylori."
RL Nature 388:539-547(1997).
CC
CC -1- SIMILARITY: BELONGS TO THE GTP1 / OBG FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000548; AAD07372.1; -
CC TIGR; HP0303; -
CC
CC InterPro: IPR000765; GTP1_OBG.
CC Pfam: PF01018; GTP1_OBG; 1.
CC PRINTS; PR00326; GTP_OBG.
CC PROSITE; PS00905; GTP1_OBG; 1.
CC KW Hypothetical protein; GTP-binding; Complete proteome.
CC FT NP_BIND 163 170 GTP (BY SIMILARITY).
CC FT NP_BIND 210 214 GTP (BY SIMILARITY).
CC FT NP_BIND 279 282 GTP (BY SIMILARITY).
CC SEQUENCE 360 AA; 38707 MW; 74FC579D99643DAC CRC64;
CC
CC
CC Query Match 46.1%; Score 41; DB 1; Length 360;
CC Best Local Similarity 53.8%; Pred. No. 6.7;
CC Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 5 VSFRRSYSLRGG 17
CC DB 20 VSFRREKFKVIGK 32
CC
CC
CC RESULT 11
CC Y389_BUCAI
CC ID Y389_BUCAI STANDARD; PRT; 334 AA.
CC AC P57469; Q9L4J5;
CC DT 20-AUG-2001 (Rel. 40, Created)
CC DT 20-AUG-2001 (Rel. 40, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE HYPOTHETICAL GTP-BINDING PROTEIN BU389.
CC GN BU389.
CC OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
CC symbiotic bacterium).
CC OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
CC OX NCBI_TaxID=118099;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=TOKYO 1998;
CC RX MEDLINE=20445173; PubMed=10993077;
CC RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
CC RT "Genome sequence of the endocellular bacterial symbiont of aphids
CC Buchnera sp. APS.";
CC FT Nature 407:81-86(2000).
CC RN [2]
CC RP SEQUENCE OF 229-334 FROM N.A.
CC RC MEDLINE=20245558; PubMed=10781569;
CC RA Jimenez N., Gonzalez-Candelas F., Silva F.J.;
CC RT "Prephenate dehydratase from the aphid endosymbiont (Buchnera)
CC displays changes in the regulatory domain that suggest its
CC desensitization to inhibition by phenylalanine.";
CC RN J. Bacteriol. 182:2967-2969(2000).
CC
CC -1- SIMILARITY: BELONGS TO THE GTP1 / OBG FAMILY. STRONG, TO
CC E.COLI YHBZ OR H.INFLUENZAE HI0877.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; AP001119; BAB13092.1; -
CC DR EMBL; AJ239043; CAB90993.1; -
CC InterPro: IPR000765; GTP1_OBG.
CC DR Pfam: PF01018; GTP1_OBG; 1.
CC PRINTS; PR00326; GTP_OBG.
CC DR PROSITE; PS00905; GTP1_OBG; FALSE_NEG.
CC KW Hypothetical protein; GTP-binding; Complete proteome.
CC FT NP_BIND 166 173 GTP (BY SIMILARITY).
CC FT NP_BIND 213 217 GTP (BY SIMILARITY).
CC FT NP_BIND 283 286 GTP (BY SIMILARITY).
CC FT CONFLICT 293 293 L -> P (IN REF. 2).
CC SEQUENCE 334 AA; 37248 MW; 54F7A07404FA7E85 CRC64;
CC
CC
CC Query Match 44.9%; Score 40; DB 1; Length 334;
CC Best Local Similarity 61.5%; Pred. No. 9.4;
CC Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
CC
CC QY 5 VSFRRSYSLRGG 17
CC DB 21 VSFRREKYIPKGG 33
CC
CC
CC RESULT 12
CC INSI_ECOLI
CC ID INSI_ECOLI STANDARD; PRT; 383 AA.
CC AC P37246; P77341;
CC DT 01-OCT-1994 (Rel. 30, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE TRANSPOSE INSI FOR INSERTION SEQUENCE ELEMENT IS30B/C/D.
CC GN (INSII OR B0256) AND (INSI2 OR B1404) AND (INSI3 OR B4284).
CC OS Escherichia coli.
CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC OX NCBI_TaxID=562;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=K12;
CC RX MEDLINE=85027168; PubMed=6092059;
CC RA Dalrymple B., Caspers P., Arber W.;
CC RT "Nucleotide sequence of the prokaryotic mobile genetic element IS30.";
CC RL EMBL J. 3:2145-2149(1984).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=K12 / MG1655;
CC RX MEDLINE=95334362; PubMed=7610040;
CC RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
CC RA Blattner F.R.;
CC RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
CC RT region from 92.8 through 100 minutes.";
CC RL Nucleic Acids Res. 23:2105-2119(1995).
CC RN [3]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=K12 / MG1655;
CC RX MEDLINE=97426617; PubMed=9278503;
CC RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
CC RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
CC RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
CC RA Mau B., Shao Y.;
CC RT "The complete genome sequence of Escherichia coli K-12.";
CC RL Science 277:1453-1474(1997).
CC RN [4]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=K12 / W3110;
CC RA Takenoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
CC RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
CC RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
CC RT "Systematic sequencing of the Escherichia coli genome: analysis of the
CC RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
CC RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC RN [5]

```

RP SEQUENCE FROM N.A.  
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,  
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,  
RA Davis R.W.,  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [16]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,  
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,  
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubraman S.,  
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,  
RA Yamamoto Y., Horluchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
CC -!- FUNCTION: REQUIRED FOR THE TRANSDUCTION OF THE INSERTION ELEMENT  
(POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE IS30 FAMILY OF TRANSPOSASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X00792; CAA25376.1; -;  
DR EMBL; U14003; AAA97180.1; -;  
DR EMBL; AE000133; AAC73359.1; -;  
DR EMBL; AE000237; AAC74486.1; -;  
DR EMBL; AE000499; AAC77240.1; -;  
DR EMBL; DB3536; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; U70214; AAB08675.1; -;  
DR EMBL; D90778; BAA15014.1; -;  
DR EMBL; D90779; BAA15020.1; -;  
DR PIR; S28740; S28740.  
DR Ecogene; EG40009; insl.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR001598; Transposase\_is30.  
DR Pfam; PF00665; rve; 1.  
DR ProDom; PD002997; Transposase\_is30; 1.  
DR ProSITE; PS01043; TRANSPOSASE\_is30; 1.  
KW Transposition; Transposable element; DNA-binding; DNA recombination;  
KW Complete proteome.  
FT VARIANT 261 261 L -> V (IN B0256).  
FT CONFLICT 5 5 F -> I (IN REF. 4).  
FT CONFLICT 27 27 A -> T (IN REF. 4).  
SQ SEQUENCE 383 AA; 44281 MW; 6AADF2CD604B0F83 CRC64;  
  
Query Match 44.9%; Score 40; DB 1; Length 383;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
QY 3 VSVFRRRSYSLRGR 18  
::: |||: ||| ||  
Db 184 LNIQHLRRSHSLRGR 199  
  
RESULT 13  
YHBZ\_ECOLI  
ID YHBZ\_ECOLI STANDARD; PRT; 390 AA.  
AC P42641;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 43.3 KDA GTP-BINDING PROTEIN IN DABC-RPMA INTERGENIC

DE REGION.  
GN YHBZ OR B3183.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1234-1238(1997).  
CC -!- SIMILARITY: BELONGS TO THE GTP1 / OBG FAMILY. STRONG, TO  
H.INFLUENZAE HI0877.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U18997; AAA57984.1; -;  
DR EMBL; AS000399; AAC76215.1; -;  
DR Ecogene; EG12795; yhbZ.  
DR InterPro; IPR000765; GTP1\_OBG.  
DR Pfam; PF01018; GTP1\_OBG; 1.  
DR PRINTS; PR00326; GTP1\_OBG.  
DR PROSITE; PS00905; GTP1\_OBG; 1.  
KW Hypothetical protein; GTP-binding; Complete proteome.  
FT NP\_BIND 166 173 GTP (BY SIMILARITY).  
FT NP\_BIND 213 217 GTP (BY SIMILARITY).  
FT NP\_BIND 283 286 GTP (BY SIMILARITY).  
SQ SEQUENCE 390 AA; 43286 MW; 3A6EBF56F24B7C47 CRC64;  
  
Query Match 44.9%; Score 40; DB 1; Length 390;  
Best Local Similarity 61.5%; Pred. No. 11;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 5 VSPFRRSYSLRGG 17  
||||| |:  
Db 21 VSPFRKYIPKGG 33  
  
RESULT 14  
PEES\_PSEAE  
ID PEES\_PSEAE STANDARD; PRT; 446 AA.  
AC Q04804;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SENSOR PROTEIN PFES (EC 2.7.3.-).  
GN PFES OR PA2687.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PAO;  
RX MEDLINE=93368425; PubMed=8361354;  
RA Dean C.R., Poole K.;  
RT "Expression of the ferric enterobactin receptor (pfeA) of Pseudomonas  
aeruginosa: involvement of a two-component regulatory system.";  
RL Mol. Microbiol. 8:1093-1103(1993).  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Iori S., Olson M.V.,  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM PFER/PFES.  
 CC -!- MAY ACTIVATE PFER BY PHOSPHORYLATION  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE  
 CC KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L07739; AA25930.1; .  
 DR EMBL; AE004696; AAG06075.1; .  
 DR InterPro; IPR000410; Bctrl\_sensord.  
 DR InterPro; IPR000658; DUF5.  
 DR InterPro; IPR003660; HAMP.  
 DR InterPro; IPR003594; HATPase\_c.  
 DR InterPro; IPR003661; His\_kinA.  
 DR Pfam; PF00672; DUF5; 1.  
 DR Pfam; PF02518; HATPase\_c; 1.  
 DR Pfam; PF00512; signal; 2.  
 DR SMART; SM00304; HAMP; 1.  
 DR SMART; SM00387; HATPase\_c; 1.  
 DR SMART; SM00388; Hiska; 1.  
 KW Sensory transduction; Transferase; Kinase; Phosphorylation;  
 KW Transmembrane; Inner membrane; Complete proteome.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 30  
 FT DOMAIN 31 155  
 FT TRANSMEM 156 176  
 FT DOMAIN 177 446  
 FT MOD\_RES 244 244  
 FT CONFLICT 304 304  
 FT A -> E (IN REF. 1).  
 SQ SEQUENCE 446 AA; 50539 MW; DIDA4FED1C222547 CRC64;

Query Match 44.98; Score 40; DB 1; Length 446;  
 Best Local Similarity 41.28; Pred. No. 13;  
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 Qy 2 GVSVSFRRSRLRGR 18  
 I : : : I : I : I :  
 Db 410 GLGLAIRARTELQGR 426

RESULT 15  
 CMGA\_HUMAN STANDARD; PRT; 457 AA.  
 AC P10645;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CHROMOGANIN A PRECURSOR (CGA) (PITUITARY SECRETORY PROTEIN I) (SP-I)  
 DE [CONTAINS: VASOSTATIN I; VASOSTATIN II; EA-92; ES-43; PANCREASTATIN;  
 DE SS-18; WA-8; WE-14; LF-19; AL-11; GV-19; GR-44; ER-37].  
 GN CHGA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86298816; PubMed=3403545;  
 RA Helman L.J., Ahn T.G., Levine M.A., Allison A., Cohen P.S.,  
 RA Cooper M.J., Cohn D.V., Israel M.A.;  
 RT "Molecular cloning and primary structure of human chromogranin A  
 RT (secretory protein I) cDNA.";  
 RL J. Biol. Chem. 263:11559-11563(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88059106; PubMed=2445752;  
 RA Konecki D.S., Benedum U.M., Gerdes H.H., Huttner W.B.;  
 RT "The primary structure of human chromogranin A and pancreastatin.";  
 RL J. Biol. Chem. 262:17026-17030(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94165095; PubMed=8120054;  
 RA Moulind A.J., Bevan S., White J.H., Hendy G.N.;  
 RT "Human chromogranin A gene. Molecular cloning, structural analysis,  
 RT and neuroendocrine cell-specific expression.";  
 RL J. Biol. Chem. 269:6918-6926(1994).  
 RN [4]  
 RP REVISIONS TO 384-397.  
 RA Moulind A.J., Bevan S., White J.H., Hendy G.N.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 19-46.  
 RC TISSUE=Adrenal gland;  
 RX MEDLINE=86206941; PubMed=3704195;  
 RA Wilson B.S., Phan S.H., Lloyd R.V.;  
 RT "Chromogranin from normal human adrenal glands: purification by  
 RT monoclonal antibody affinity chromatography and partial N-terminal  
 RT amino acid sequence.";  
 RL Regul. Pept. 13:207-233(1986).  
 RN [6]  
 RP SEQUENCE OF 134-319.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=90336639; PubMed=2165909;  
 RA Tamamura H., Ohta M., Yoshizawa K., Ono Y., Funakoshi A.,  
 RA Miyasaka K., Tateishi K., Jimi A., Yajima H., Fujii N., Funakoshi S.;  
 RT "Isolation and characterization of a tumor-derived human protein  
 RT related to chromogranin A and its in vitro conversion to human  
 RT pancreastatin-48.";  
 RL Eur. J. Biochem. 191:33-39(1990).  
 RN [7]  
 RP SEQUENCE OF 291-319.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=88137586; PubMed=2830133;  
 RA Sekiya K., Gbatei M.A., Minamino N., Bretherton-Watt D., Matsuo H.,  
 RA Bloom S.R.;  
 RT "Isolation of human pancreastatin fragment containing the active  
 RT sequence from a glucagonoma.";  
 RL FEBS Lett. 228:153-156(1988).  
 RN [8]  
 RP SEQUENCE OF 342-355.  
 RX MEDLINE=92249591; PubMed=1577173;  
 RA Curry W.J., Shaw C., Johnston C.F., Thim L., Buchanan K.D.;  
 RT "Isolation and primary structure of a novel chromogranin A-derived  
 RT peptide, WE-14, from a human midgut carcinoid tumour.";  
 RL FEBS Lett. 301:319-321(1992).  
 RN [9]  
 RP SEQUENCE OF DERIVED PEPTIDES.  
 RA Orr D.F., Salmon A.L., Johnsen A.H., Chalk R., Buchanan K.D.,  
 RA Shaw C.;  
 RT "The spectrum of endogenous human chromogranin A-derived peptides  
 RT identified using a modified proteomic strategy.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1999).  
 RN [10]  
 RP CARBOHYDRATE-LINKAGE SITES, AND PHOSPHORYLATION.  
 RX PubMed=9852066;





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:35 ; Search time 126.85 Seconds  
(without alignments)  
10.809 Million cell updates/sec

Title: US-09-485-571-17  
Perfect score: 89  
Sequence: 1 RGVSVSFRRSYSLRGGR 18  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	49.4	192	2 S42988	viral infectivity
2	43	48.3	449	1 A41520	chromogranin A pre
3	41	46.1	188	1 ASLJZR	vif protein - huma
4	41	46.1	192	1 ASLJND	vif protein - huma
5	41	46.1	192	2 S42999	viral infectivity
6	41	46.1	192	2 S42966	viral infectivity
7	41	46.1	192	2 S43000	viral infectivity
8	41	46.1	192	2 S42997	viral infectivity
9	41	46.1	192	2 S42998	viral infectivity
10	41	46.1	192	2 S42968	viral infectivity
11	41	46.1	192	2 S42980	viral infectivity
12	41	46.1	192	2 S42959	viral infectivity
13	41	46.1	192	2 S42945	viral infectivity
14	41	46.1	192	2 S43004	viral infectivity
15	41	46.1	192	2 S42974	viral infectivity
16	41	46.1	192	2 S42958	viral infectivity
17	41	46.1	192	2 S42979	viral infectivity
18	41	46.1	192	2 S42961	viral infectivity
19	41	46.1	192	2 S42965	viral infectivity
20	41	46.1	192	2 S42955	viral infectivity
21	41	46.1	192	2 S42940	viral infectivity
22	41	46.1	192	2 S54379	vif protein - huma
23	41	46.1	192	2 S42954	viral infectivity
24	41	46.1	192	2 S42953	viral infectivity
25	41	46.1	202	2 S50116	ribosomal protein
26	41	46.1	289	2 D69152	hypothetical prote
27	41	46.1	360	1 G64557	GTP-binding protei
28	41	46.1	360	2 B71952	hypothetical prote
29	41	46.1	445	2 G83529	two-component sens

30	41	46.1	849	2 E86306	Similar to tuftell
31	41	46.1	874	2 E75278	hypothetical prote
32	41	46.1	1513	2 T44045	hypothetical prote
33	41	46.1	1520	2 T44231	hypothetical prote
34	41	46.1	2352	2 T06077	splicing factor PR
35	41	46.1	2359	2 B96832	hypothetical prote
36	40.5	45.5	645	2 D85844	unknown protein en
37	40.5	45.5	645	2 A85629	hypothetical prote
38	40.5	45.5	645	2 D85680	unknown protein en
39	40	44.9	192	2 S42967	viral infectivity
40	40	44.9	251	2 A85911	IS30 transposase I
41	40	44.9	334	2 D84975	hypothetical prote
42	40	44.9	376	2 C71286	probable GTP-bind
43	40	44.9	383	1 F65241	transposase - Esch
44	40	44.9	390	1 A65109	GTP-binding protei
45	40	44.9	390	2 A85982	probable GTP-bind

ALIGNMENTS

RESULT 1

S42988

viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999

C:Accession: S42988

R;Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994

A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940

A:Accession: S42988

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-192 <WIE>

A:Cross-references: EMBL:Z30679; NID:g459595; PIDN:CAA83159.1; PID:g459596  
C:Superfamily: AIDS vif protein

Query Match 49.4%; Score 44; DB 2; Length 192;  
Best Local Similarity 47.1%; Pred. No. 4.6;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGVSVSFRRSYSLRGGR 17

Db 83 QGVSVIEWKRRYSTQVG 99

RESULT 2

A41520

chromogranin A precursor [validated] - bovine  
N:Alternate names: pituitary secretory protein I; secretory protein I  
N:Contains: chromostatin; pancreastatin  
C:Species: Bos primigenius taurus (cattle)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Aug-2000

C:Accession: A41520; A24175; A60306; A61114; S15847; S39016; I46008; S38976

R;Iacangelo, A.L.; Grimes, M.; Eiden, L.E.

Mol. Endocrinol. 5, 1651-1660, 1991

A:Title: The bovine chromogranin A gene: structural basis for hormone regulation an

A:Reference number: A41520; MUID:92140395

A:Accession: A41520

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-449 <TAC>

A:Cross-references: GB:S79270; NID:g244423; PIDN:AAB21297.1; PID:g244424  
R;Ahm, T.G.; Cohn, D.V.; Gorr, S.U.; Ornstein, D.L.; Kashdan, M.A.; Levine, M.A.

Proc. Natl. Acad. Sci. U.S.A. 84, 5043-5047, 1987

A:Title: Primary structure of bovine pituitary secretory protein I (chromogranin A)

A:Reference number: A28033; MUID:87260925

A:Accession: A28033

A:Molecule type: mRNA

A:Residues: 1-111, 'T', 113-190, 'Y', 192-253, 'P', 255-378, 'R', 380-449 <AHN>  
A:Cross-references: GB:M16971; NID:g163727; PIDN:AAA30765.1; PID:g163728

A:Note: the authors translated the codon CGG for residue 391 as Gln  
 R:Iacangelo, A.; Affolter, H.U.; Eiden, L.E.; Herbert, E.; Grimes, M.  
 Nature 323, 82-86, 1986  
 A:Title: Bovine chromogranin A sequence and distribution of its messenger RNA in endocrine  
 A:Reference number: A24175; MUID:86311345  
 A:Accession: A24175  
 A:Molecule type: mRNA  
 A:Residues: 1-153, 'PO', 156-158, 'R', 160-190, 'Y', 192-253, 'P', 255-449 <IA2>  
 A:Cross-references: GB:X04298; NID:g217; PIDN:CAA27841.1; PID:g218  
 R:Nakano, I.; Funakoshi, A.; Miyasaka, K.; Ishida, K.; Makk, G.; Angwin, P.; Chang, D.;  
 Regul. Pept. 25, 207-213, 1989  
 A:Title: Isolation and characterization of bovine pancreastatin.  
 A:Reference number: A60306; MUID:89331945  
 A:Accession: A60306  
 A:Molecule type: protein  
 A:Residues: 266-312 <NAK>  
 R:Barbosa, J.A.; Gill, B.M.; Takiyuddin, M.A.; O'Connor, D.T.  
 Endocrinology 128, 174-190, 1991  
 A:Title: Chromogranin A: posttranslational modifications in secretory granules.  
 A:Reference number: A61114; MUID:91099142  
 A:Accession: A61114  
 A:Molecule type: protein  
 A:Residues: 19-34, 'X', 36-38; 97-111; 134-139 <BA2>  
 R:Watkinson, A.; Joensson, A.C.; Davison, M.; Young, J.; Lee, C.M.; Moore, S.; Dockray, C.  
 Biochem. J. 276, 471-479, 1991  
 A:Title: Heterogeneity of chromogranin A-derived peptides in bovine gut, pancreas and adrenal  
 A:Reference number: S15847; MUID:91264803  
 A:Accession: S15847  
 A:Molecule type: protein  
 A:Residues: 266-310, 'H', 312-318, 'K', 320-331 <WA2>  
 A:Note: 311-Arg and 319-Glu were also found  
 R:Watkinson, A.; Rogers, M.; Dockray, G.J.  
 Biochem. J. 295, 649-654, 1993  
 A:Title: Post-translational processing of chromogranin A: differential distribution of  
 A:Reference number: S39016; MUID:94059013  
 A:Accession: S39016  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 303-331 <WAT>  
 R:Benedict, U.M.; Baeuerle, P.A.; Konecki, D.S.; Frank, R.; Powell, J.; Mallet, J.; Hutter  
 EMBO J. 5, 1495-1502, 1986  
 A:Title: The primary structure of bovine chromogranin A: a representative of a class of  
 A:Reference number: I46008; MUID:86300648  
 A:Accession: I46008  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-135, 'S', 137-190, 'Y', 192-253, 'P', 255-310, 'H', 312-318, 'K', 320-390, 'Q', 392-444  
 A:Cross-references: EMBL:X04012; NID:g197; PIDN:CAA27636.1; PID:g198  
 R:Yoo, S.H.; Ferretti, J.A.  
 FEBS Lett. 334, 373-377, 1993  
 A:Title: Nature of the pH-induced conformational changes and exposure of the C-terminal  
 A:Reference number: S38976; MUID:94063061  
 A:Accession: S38976  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 19-26; 266-272 <YOO>  
 A:Comment: Chromogranin A is the major protein of bovine chromaggin granules.  
 C:Comment: Chromogranin activity has been demonstrated from proteolytic fragments of chr  
 C:Genetics:  
 A:Introns: 16/1; 31/3; 63/1; 86/1; 119/1; 266/1; 422/3  
 C:Superfamily: chromogranin A  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-449/Product: chromogranin A #status experimental <MAT>  
 F:142-161/Product: chromogranin A #status predicted <MAT3>  
 F:266-312/Product: pancreastatin #status experimental <MAT2>  
 F:35-56/Disulfide bonds: #status predicted

Query Match 48.3%; Score 43; DB 1; Length 449;  
 Best Local Similarity 50.0%; Pred. No. 16;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYSLRG 16  
 Db 362 RSMRLSFRAGYGFGRG 377

RESULT 3  
 ASLJZR  
 vif protein - human immunodeficiency virus 2r-6  
 N:Alternate names: orf-Q protein; sor protein  
 C:Species: human immunodeficiency virus 2r-6  
 C>Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 16-Jul-1999  
 C:Accession: B26192  
 R:Srinivasan, A.; Anand, R.; York, D.; Ranganathan, P.; Feorino, P.; Schochetman, G.  
 Gene 52, 71-82, 1987  
 A:Title: Molecular characterization of human immunodeficiency virus from Zaire: nucle  
 A:Reference number: A26192; MUID:87248097  
 A:Accession: B26192  
 A:Molecule type: DNA  
 A:Residues: 1-188 <SRT>  
 A:Cross-references: GB:K03458; GB:M16322; NID:g329398; PIDN:AAAM5379.1; PID:g329402  
 C:Genetics:  
 A:Gene: vif  
 C:Superfamily: AIDS vif protein  
 C:Keywords: AIDS; immunodeficiency

Query Match 46.1%; Score 41; DB 1; Length 188;  
 Best Local Similarity 53.8%; Pred. No. 14;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYS 13  
 Db 83 QGVSIWKRKRRYS 95

RESULT 4  
 ASLJND  
 vif protein - human immunodeficiency virus type 1 (isolate NDK)  
 N:Alternate names: orf-Q protein; sor protein  
 C:Species: human immunodeficiency virus type 1, HIV-1  
 A:Note: host Homo sapiens (man)  
 C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999  
 C:Accession: JQ0069  
 R:Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe,  
 Gene 81, 275-284, 1989  
 A:Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human im  
 A:Reference number: JQ0065; MUID:90034200  
 A:Accession: JQ0069  
 A:Molecule type: DNA  
 A:Residues: 1-192 <SPI>  
 A:Cross-references: GB:M27323; NID:g328154; PIDN:AAA44870.1; PID:g328159  
 C:Genetics:  
 A:Gene: vif  
 C:Superfamily: AIDS vif protein  
 C:Keywords: AIDS; immunodeficiency

Query Match 46.1%; Score 41; DB 1; Length 192;  
 Best Local Similarity 53.8%; Pred. No. 15;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYS 13  
 Db 83 QGVSIWKRKRRYS 95

RESULT 5  
 S42999  
 viral infectivity factor vif - human immunodeficiency virus type 1  
 C:Species: human immunodeficiency virus type 1, HIV-1  
 C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
 C:Accession: S42999  
 R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.

submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42999  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30690; NID:g459617; PIDN:CAA83170.1; PID:g459618  
C:Superfamily: AIDS vif protein

Query Match 46.1%; Score 41; DB 2; Length 192;  
Best Local Similarity 53.8%; Pred. No. 15;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|:|: :|:|:|  
Db 83 QGVSIWRRRRYS 95

RESULT 6  
S42966  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42966  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42966  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30627; NID:g459496; PIDN:CAA83104.1; PID:g459497  
C:Superfamily: AIDS vif protein

Query Match 46.1%; Score 41; DB 2; Length 192;  
Best Local Similarity 53.8%; Pred. No. 15;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|:|: :|:|:|  
Db 83 QGVSIWRRRRYS 95

RESULT 7  
S43000  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S43000  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S43000  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30691; NID:g459619; PIDN:CAA83171.1; PID:g459620  
C:Superfamily: AIDS vif protein

Query Match 46.1%; Score 41; DB 2; Length 192;  
Best Local Similarity 53.8%; Pred. No. 15;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|:|: :|:|:|  
Db 83 QGVSIWRRRRYS 95

RESULT 8  
S42997

viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42997  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42997  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30688; NID:g459613; PIDN:CAA83168.1; PID:g459614  
C:Superfamily: AIDS vif protein

Query Match 46.1%; Score 41; DB 2; Length 192;  
Best Local Similarity 53.8%; Pred. No. 15;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|:|: :|:|:|  
Db 83 QGVSIWRRRRYS 95

RESULT 9  
S42998  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42998  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42998  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30689; NID:g459615; PIDN:CAA83169.1; PID:g459616  
C:Superfamily: AIDS vif protein

Query Match 46.1%; Score 41; DB 2; Length 192;  
Best Local Similarity 53.8%; Pred. No. 15;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYS 13  
:|:|: :|:|:|  
Db 83 QGVSIWRRRRYS 95

RESULT 10  
S42968  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42968  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42968  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30629; NID:g459500; PIDN:CAA83106.1; PID:g459501  
C:Superfamily: AIDS vif protein

Query Match  
Best Local Similarity 46.1%; Score 41; DB 2; Length 192;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGVSVSFRRRSYS 13  
:|:|: |:|:| |  
Db 83 QGVSIWKRKRRYS 95

RESULT 11  
S42980  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42980  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42980  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30641; NID:g459524; PIDN:CAA83118.1; PID:g459525  
C:Superfamily: AIDS vif protein

Query Match  
Best Local Similarity 46.1%; Score 41; DB 2; Length 192;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGVSVSFRRRSYS 13  
:|:|: |:|:| |  
Db 83 QGVSIWKRKRRYS 95

RESULT 12  
S42959  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42959  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42959  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30620; NID:g459482; PIDN:CAA83097.1; PID:g459483  
C:Superfamily: AIDS vif protein

Query Match  
Best Local Similarity 46.1%; Score 41; DB 2; Length 192;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGVSVSFRRRSYS 13  
:|:|: |:|:| |  
Db 83 QGVSIWKRKRRYS 95

RESULT 13  
S42945  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42945  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940

A:Accession: S42945  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30606; NID:g459456; PIDN:CAA83083.1; PID:g459457  
C:Superfamily: AIDS vif protein

Query Match  
Best Local Similarity 46.1%; Score 41; DB 2; Length 192;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGVSVSFRRRSYS 13  
:|:|: |:|:| |  
Db 83 QGVSIWKRKRRYS 95

RESULT 14  
S43004  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S43004  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S43004  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30695; NID:g459627; PIDN:CAA83175.1; PID:g459628  
C:Superfamily: AIDS vif protein

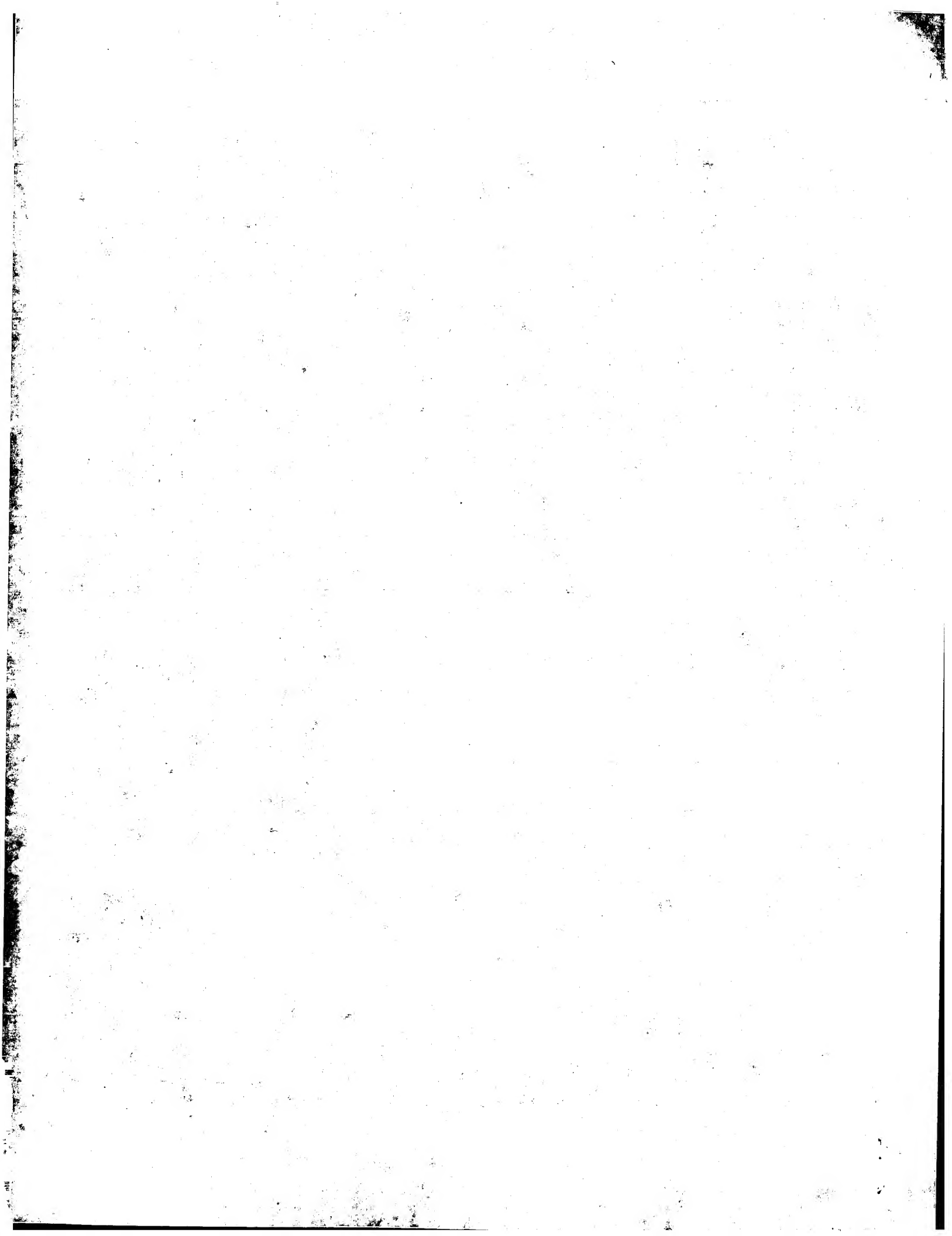
Query Match  
Best Local Similarity 46.1%; Score 41; DB 2; Length 192;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGVSVSFRRRSYS 13  
:|:|: |:|:| |  
Db 83 QGVSIWKRKRRYS 95

RESULT 15  
S42974  
viral infectivity factor vif - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C>Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S42974  
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.  
submitted to the EMBL Data Library, March 1994  
A:Description: In vivo genetic variability of the HIV-1 gene.  
A:Reference number: S42940  
A:Accession: S42974  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <WIE>  
A:Cross-references: EMBL:Z30635; NID:g459512; PIDN:CAA83112.1; PID:g459513  
C:Superfamily: AIDS vif protein

Query Match  
Best Local Similarity 46.1%; Score 41; DB 2; Length 192;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 RGVSVSFRRRSYS 13  
:|:|: |:|:| |  
Db 83 QGVSIWKRKRRYS 95

Job time: 554 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:21 ; Search time 106.12 Seconds  
(without alignments)  
3.817 Million cell updates/sec

Title: US-09-485-571-17

Perfect score: 89

Sequence: 1 RGVSVSPRRRSYSLRGGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*  
1: /cgn2.6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2.6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2.6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2.6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2.6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	77.5	18	2	US-08-752-852A-15
2	54.5	61.2	17	2	US-08-752-852A-14
3	47	52.8	15	2	US-08-752-852A-10
4	47	52.8	15	2	US-08-752-852A-11
5	47	52.8	15	2	US-08-752-852A-16
6	41	46.1	90	4	US-08-679-493A-93
7	40	44.9	18	1	US-08-182-483A-28
8	40	44.9	18	1	US-08-243-879A-27
9	40	44.9	18	1	US-08-499-523-48
10	40	44.9	18	4	US-09-128-345-48
11	39	43.8	27	3	US-08-886-269-7
12	39	43.8	27	3	US-09-167-647-7
13	39	43.8	165	3	US-08-799-149C-5
14	39	43.8	182	4	US-08-893-764-2
15	39	43.8	236	3	US-08-886-269-1
16	39	43.8	236	3	US-08-886-269-2
17	39	43.8	236	3	US-09-167-647-1
18	39	43.8	236	3	US-09-167-647-2
19	38	42.7	15	2	US-08-752-852A-17
20	38	42.7	192	4	US-09-124-900-4
21	38	42.7	202	4	US-09-342-084-10
22	38	42.7	203	3	US-08-463-210-10
23	37	41.6	335	2	US-09-014-969-15
24	36.5	41.0	355	2	US-08-641-314C-2
25	36	40.4	13	2	US-08-752-852A-12
26	36	40.4	101	4	US-08-679-493A-26
27	36	40.4	306	1	US-08-233-788A-50

28 36 40.4 621 3 US-09-026-343-7 Sequence 7, Appl  
29 35 39.3 146 3 US-08-966-317-4 Sequence 4, Appl  
30 35 39.3 353 3 US-08-803-603-3 Sequence 3, Appl  
31 35 39.3 488 1 US-07-794-393-2 Sequence 2, Appl  
32 35 39.3 488 1 US-08-001-711-2 Sequence 2, Appl  
33 35 39.3 488 3 US-08-704-711A-22 Sequence 22, Appl  
34 35 39.3 489 4 US-08-448-489-11 Sequence 11, Appl  
35 35 39.3 494 4 US-09-413-814-88 Sequence 88, Appl  
36 35 39.3 503 3 US-08-803-603-1 Sequence 1, Appl  
37 35 39.3 503 3 US-08-369-822C-8 Sequence 8, Appl  
38 35 39.3 503 3 US-08-779-764A-26 Sequence 26, Appl  
39 35 39.3 503 3 US-08-779-764A-27 Sequence 27, Appl  
40 35 39.3 503 3 US-08-582-776C-8 Sequence 8, Appl  
41 35 39.3 503 3 US-08-434-831B-8 Sequence 8, Appl  
42 35 39.3 509 2 US-08-565-926-8 Sequence 10, Appl  
43 35 39.3 509 4 US-08-740-223A-10 Patent No. RE34,60  
44 35 39.3 522 6 RE34606-6  
45 35 39.3 564 3 US-08-851-843A-101 Sequence 101, App

#### ALIGNMENTS

RESULT 1  
US-08-752-852A-15  
; Sequence 15, Application US/08752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrer, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cotuzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Other  
; LOCATION: 1...18

OTHER INFORMATION: all amino acids are D amino acids  
US-08-752-852A-15

Query Match 77.5%; Score 69; DB 2; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.2e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYSLRGR 18  
||| | |||| | ||||  
DB 1 RGVCFRRRCYCLRGR 18

## RESULT 2

US-08-752-852A-14  
; Sequence 14, Application US/08752852A  
; Patent No. 5994306

## GENERAL INFORMATION:

APPLICANT: Chang, Conway  
APPLICANT: Gu, Chee-Liang  
APPLICANT: Chen, Jie  
APPLICANT: Steinberg, Deborah  
APPLICANT: Lehrer, Robert  
APPLICANT: Harwig, Sylvia

TITLE OF INVENTION: FINE-TUNED PROTEGRINS

NUMBER OF SEQUENCES: 242

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/752,852A

FILING DATE: 21-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 8067-034-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-9741

TELEX: 66141

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 17 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-752-852A-14

Query Match 61.2%; Score 54.5; DB 2; Length 17;  
Best Local Similarity 72.2%; Pred. No. 0.0032;  
Matches 13; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 RGVSVFRRRSYSLRGR 18  
||| | |||| | ||||  
DB 1 RGVCFRRRCY-CRGR 17

## RESULT 3

US-08-752-852A-10  
; Sequence 10, Application US/08752852A  
; Patent No. 5994306

## GENERAL INFORMATION:

APPLICANT: Chang, Conway

APPLICANT: Gu, Chee-Liang

APPLICANT: Chen, Jie

APPLICANT: Steinberg, Deborah

APPLICANT: Lehrer, Robert

APPLICANT: Harwig, Sylvia

TITLE OF INVENTION: FINE-TUNED PROTEGRINS

NUMBER OF SEQUENCES: 242

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/752,852A

FILING DATE: 21-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 8067-034-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-9741

TELEX: 66141

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-752-852A-10

Query Match 52.8%; Score 47; DB 2; Length 15;  
Best Local Similarity 71.4%; Pred. No. 0.053;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYSL 14  
||| | |||| | ||||  
DB 1 RGVCFRRRCYCL 14

## RESULT 4

US-08-752-852A-11  
; Sequence 11, Application US/08752852A  
; Patent No. 5994306

## GENERAL INFORMATION:

APPLICANT: Chang, Conway

APPLICANT: Gu, Chee-Liang

APPLICANT: Chen, Jie

APPLICANT: Steinberg, Deborah

APPLICANT: Lehrer, Robert

APPLICANT: Harwig, Sylvia

TITLE OF INVENTION: FINE-TUNED PROTEGRINS

NUMBER OF SEQUENCES: 242

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP

STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,852A  
FILING DATE: 21-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-034-999  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-752-852A-11

Query Match 52.8%; Score 47; DB 2; Length 15;  
Best Local Similarity 71.4%; Pred. No. 0.053;  
Matches 10; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 RGVSVSFRRRSYSL 14  
||| | |||| | |  
Db 1 RGVCVCFRRRCYCL 14

RESULT 5  
US-08-752-852A-16  
; Sequence 16, Application US/08/752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrer, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752,852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-034-999  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-752-852A-16

Query Match 52.8%; Score 47; DB 2; Length 15;  
Best Local Similarity 71.4%; Pred. No. 0.053;  
Matches 10; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

QY 1 RGVSVSFRRRSYSL 14  
||| | |||| | |  
Db 1 RGVCVCFRRRCYCL 14

RESULT 6  
US-08-679-493A-93  
; Sequence 93, Application US/08/679493A  
; Patent No. 6303295  
; GENERAL INFORMATION:  
; APPLICANT: Taylor, Ethan W.  
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS  
; FILE REFERENCE: 55-95  
; CURRENT APPLICATION NUMBER: US/08/679,493A  
; CURRENT FILING DATE: 1996-07-12  
; PRIOR APPLICATION NUMBER: 60/001203  
; PRIOR FILING DATE: 1995-07-14  
; PRIOR APPLICATION NUMBER: 60/003,112  
; PRIOR FILING DATE: 1995-09-01  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 93  
; LENGTH: 90  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-08-679-493A-93

Query Match 46.1%; Score 41; DB 4; Length 90;  
Best Local Similarity 46.7%; Pred. No. 3.9;  
Matches 7; Conservative 4; Mismatches 4; Indels 4; Gaps 0;

QY 1 RGVSVSFRRRSYSLR 15  
: | : | | | :  
Db 9 KGSSIRYRCYSIR 23

RESULT 7  
US-08-182-483A-28  
; Sequence 28, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER  
 STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20006-1812  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/182.483A  
 FILING DATE: 13-JAN-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MURASHIGE, KATE H.  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 2000-0540.21  
 TELEPHONE: (202) 887-1500  
 TELEFAX: (202) 887-0763  
 TELEX: 90-4030  
 INFORMATION FOR SEQ ID NO: 28:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 18 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-182-483A-28

Query Match 44.9%; Score 40; DB 1; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 0.99;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYSLRGR 18  
 | :||| :| :||  
 Db 1 RATRISFRRRFSVSVGR 18

RESULT 8  
 US-08-243-879A-27  
 ; Sequence 27, Application US/08243879A  
 ; Patent No. 5708145  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; TITLE OF INVENTION: A NEW PROTEGRIN  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington, DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/243,879A  
 ; FILING DATE: 17-MAY-1994  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.22  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 27:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 18 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-243-879A-27

Query Match 44.9%; Score 40; DB 1; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 0.99;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYSLRGR 18  
 | :||| :| :||  
 Db 1 RATRISFRRRFSVSVGR 18

RESULT 9  
 US-08-499-523-48  
 ; Sequence 48, Application US/08499523  
 ; Patent No. 5804558  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,523  
 ; FILING DATE: 07-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.24  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030  
 ; INFORMATION FOR SEQ ID NO: 48:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-499-523-48

Query Match 44.9%; Score 40; DB 1; Length 18;  
 Best Local Similarity 44.4%; Pred. No. 0.99;  
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGVSVFRRRSYSLRGR 18  
 | :||| :| :||  
 Db 1 RATRISFRRRFSVSVGR 18

RESULT 10  
 US-09-128-345-48  
 ; Sequence 48, Application US/09128345

; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S. I.  
; APPLICANT: KORYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-0054-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-128-345-48

Query Match 44.9%; Score 40; DB 4; Length 18;  
Best Local Similarity 44.4%; Pred. No. 0.99;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RGVSVFRRRSYSLRGR 18  
Db 1 RATRISFRRRFSVSGR 18

RESULT 11  
US-08-886-269-7  
; Sequence 7, Application US/08886269  
; Patent No. 6025148  
; GENERAL INFORMATION:  
; APPLICANT: Grieninger, Gerd  
; APPLICANT: Fu, Yiping  
; APPLICANT: Cao, Yan  
; APPLICANT: Ahadi, Mohamad Zaher  
; APPLICANT: Kudryk, Bohdan J.  
; TITLE OF INVENTION: MONOSPECIFIC ANTIBODIES AGAINST A SUBUNIT OF FIBRINOGEN  
; FILE REFERENCE: Sequences 1-7 for 454-7 CON/CPA  
; CURRENT APPLICATION NUMBER: US/08/886,269  
; CURRENT FILING DATE: 1997-07-01  
; EARLIER APPLICATION NUMBER: 08/479,755  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-08-886-269-7

Query Match 43.8%; Score 39; DB 3; Length 27;  
Best Local Similarity 62.5%; Pred. No. 2.3;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 VVSFRRRSYSLRGR 18  
Db 3 VWSFRGADYSLRAVR 18

RESULT 12  
US-09-167-647-7  
; Sequence 7, Application US/09167647  
; Patent No. 6025476  
; GENERAL INFORMATION:  
; APPLICANT: Grieninger, Gerd  
; APPLICANT: Fu, Yiping  
; APPLICANT: Cao, Yan  
; APPLICANT: Ahadi, Mohamad Zaher  
; APPLICANT: Kudryk, Bohdan J.  
; TITLE OF INVENTION: MONOSPECIFIC ANTIBODIES AGAINST A SUBUNIT OF FIBRINOGEN  
; FILE REFERENCE: Seq ID #1 - 7 454-7 DIV  
; CURRENT APPLICATION NUMBER: US/09/167,647  
; CURRENT FILING DATE: 1998-10-06  
; EARLIER APPLICATION NUMBER: 08/479,755  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-167-647-7

Query Match 43.8%; Score 39; DB 3; Length 27;  
Best Local Similarity 62.5%; Pred. No. 2.3;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 VVSFRRRSYSLRGR 18  
Db 3 VWSFRGADYSLRAVR 18

RESULT 13  
US-08-799-149C-5  
; Sequence 5, Application US/08799149C  
; Patent No. 6008195  
; GENERAL INFORMATION:  
; APPLICANT: Michael E. Selsted  
; TITLE OF INVENTION: Antimicrobial peptides and  
; TITLE OF INVENTION: Methods of Use  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/799,149C  
; FILING DATE: 14-February-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,834  
; FILING DATE: 16-February-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lisa A. Haile, Ph.D.

/ REGISTRATION NUMBER: 38,347  
/ REFERENCE/DOCKET NUMBER: 07306/009001  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 619/678-5070  
/ TELEFAX: 619/678-5099  
/ INFORMATION FOR SEQ ID NO: 5:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 165 amino acids  
/ TYPE: amino acid  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
/ FEATURE:  
/ NAME/KEY: Coding Sequence  
/ LOCATION: 36..521  
/ US-08-799-149C-5

Query Match 43.8%; Score 39; DB 3; Length 165;  
Best Local Similarity 47.1%; Pred. No. 17;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 GVSVSFRRRSYSLRGR 18  
||| | | : | |  
Db 125 GVSRGFLRSNYEVKGHR 141

## RESULT 14

US-08-893-764-2  
/ Sequence 2, Application US/08893764  
/ Patent No. 6172211  
/ GENERAL INFORMATION:

/ APPLICANT: Georgiev, Georgii P.  
/ APPLICANT: Kiselev, Sergel L.  
/ APPLICANT: Prokhorchouk, Egor B.  
/ APPLICANT: Ostermann, Elinborg  
/ TITLE OF INVENTION: Tumor Growth Inhibition- and Apoptosis-Associated  
/ TITLE OF INVENTION: Genes and Methods of Use Thereof  
/ NUMBER OF SEQUENCES: 5  
/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
/ STREET: 1100 New York Avenue, N.W., Suite 600  
/ CITY: Washington  
/ STATE: D.C.  
/ COUNTRY: U.S.A.  
/ ZIP: 20005-3934

## COMPUTER READABLE FORM:

/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/893,764  
/ FILING DATE: (Herewith)

## CLASSIFICATION: 515

/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Esmond, Robert W.  
/ REGISTRATION NUMBER: 32,893  
/ REFERENCE/DOCKET NUMBER: 0652.1630000  
/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (202) 371-2600

/ TELEFAX: (202) 371-2540

/ INFORMATION FOR SEQ ID NO: 2:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 182 amino acids

/ TYPE: amino acid

/ TOPOLOGY: linear

/ MOLECULE TYPE: protein

US-08-893-764-2

Query Match 43.8%; Score 39; DB 4; Length 182;  
Best Local Similarity 47.1%; Pred. No. 18;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 GVSVSFRRRSYSLRGR 18  
||| | | : | |  
Db 142 GVSRGFLRSNYEVKGHR 158

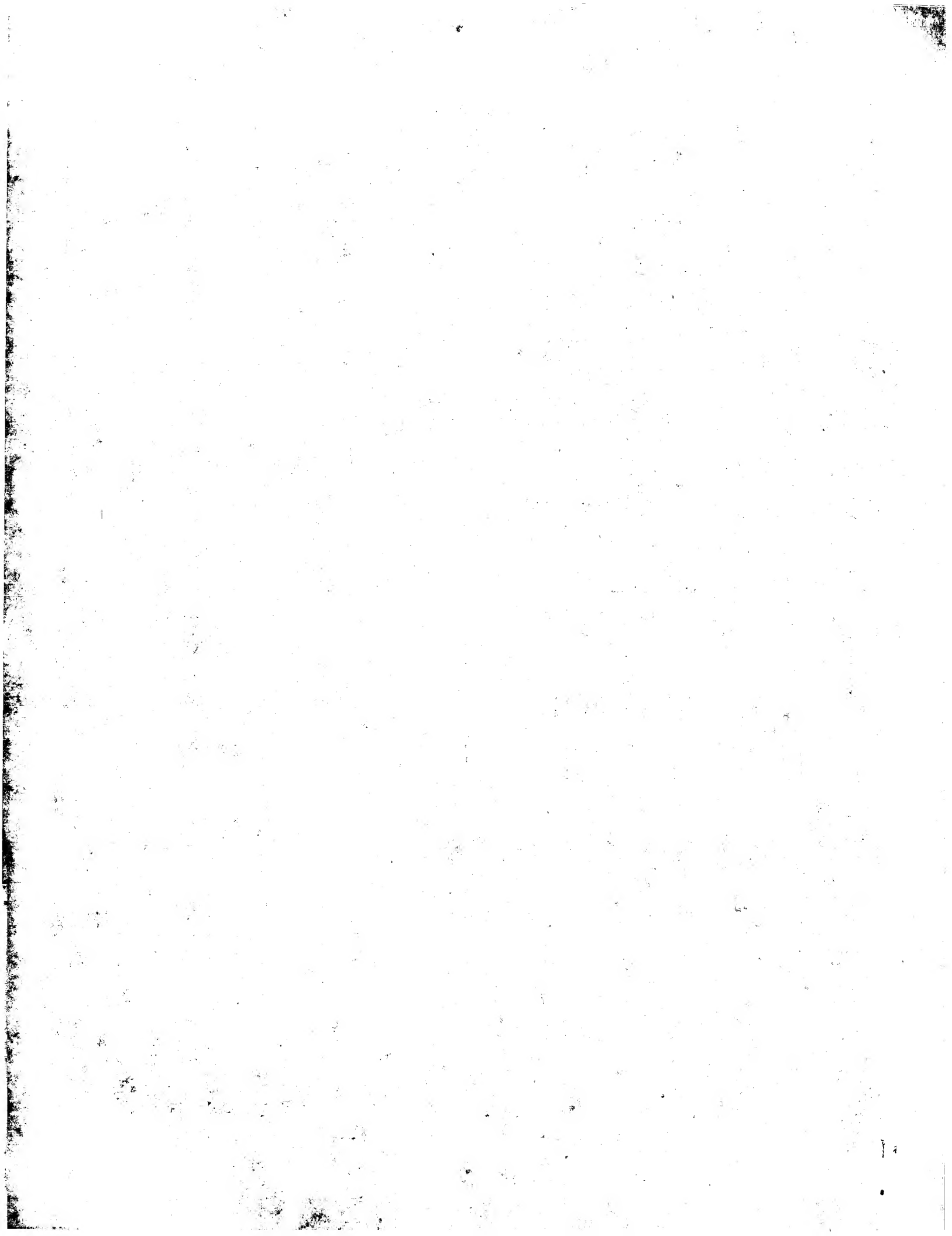
RESULT 15  
US-08-886-269-1  
/ Sequence 1, Application US/08886269  
/ Patent No. 6025148  
/ GENERAL INFORMATION:  
/ APPLICANT: Grieninger, Gerd  
/ APPLICANT: Fu, Yiping  
/ APPLICANT: Cao, Yan  
/ APPLICANT: Ahadi, Mohamad Zaher  
/ APPLICANT: Kudryk, Bohdan J.  
/ TITLE OF INVENTION: MONOSPECIFIC ANTIBODIES AGAINST A SUBUNIT OF FIBRINOGEN  
/ FILE REFERENCE: Sequences 1-7 for 454-7 CON/CPA  
/ CURRENT APPLICATION NUMBER: US/08/886,269  
/ CURRENT FILING DATE: 1997-07-01  
/ EARLIER APPLICATION NUMBER: 08/479,755  
/ EARLIER FILING DATE: 1995-06-07  
/ NUMBER OF SEQ ID NOS: 7  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 1  
/ LENGTH: 236  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
/ US-08-886-269-1

Query Match 43.8%; Score 39; DB 3; Length 236;  
Best Local Similarity 62.5%; Pred. No. 25;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 VSVSFRRSYSLRGR 18  
| | | | | | | | | |  
Db 212 VVVSFRGADYSLRAVR 227

Search completed: February 12, 2002, 12:32:22  
Job time: 450 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:30:29 ; Search time 242.57 Seconds  
(without alignments)  
5.497 Million cell updates/sec

Title: US-09-485-571-17

Perfect score: 89

Sequence: 1 RGVSFSPRRSRLRGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	18	AAW99404	Protegrin derivati
2	69	77.5	18	AAW36214	Antimicrobial prot
3	54.5	61.2	17	AAW36213	Antimicrobial prot
4	47	52.8	15	AAW36209	Antimicrobial prot
5	47	52.8	15	AAW36210	Antimicrobial prot
6	47	52.8	15	AAW36215	Antimicrobial prot
7	42	47.2	366	21 AAY77962	A. thaliana enviro
8	42	47.2	3025	22 AAB86197	HIV-1 subtype C pr
9	41	46.1	59	21 AAG25990	zea mays protein f
10	41	46.1	141	21 AAG45699	Arabidopsis thalia
11	41	46.1	142	21 AAG40714	zea mays protein f

12	41	46.1	190	20 AAY01969	HIV-1 viral infect
13	41	46.1	190	20 AAY01968	Consensus sequence
14	41	46.1	192	9 AAP81855	Sequence encoded b
15	41	46.1	192	11 AAR09302	Sequence deduced f
16	41	46.1	192	20 AAY01973	HIV-1 viral infect
17	41	46.1	192	20 AAY01974	HIV-1 viral infect
18	41	46.1	192	20 AAY01975	HIV-1 viral infect
19	41	46.1	192	20 AAY01976	HIV-1 viral infect
20	41	46.1	192	20 AAY01977	HIV-1 viral infect
21	41	46.1	192	20 AAY01981	HIV-1 viral infect
22	41	46.1	192	20 AAY01982	HIV-1 viral infect
23	41	46.1	192	20 AAY01983	HIV-1 viral infect
24	41	46.1	192	20 AAY01984	HIV-1 viral infect
25	41	46.1	192	20 AAY01985	HIV-1 viral infect
26	41	46.1	192	20 AAY01987	HIV-1 viral infect
27	41	46.1	192	20 AAY01988	HIV-1 viral infect
28	41	46.1	192	20 AAY01970	HIV-1 viral infect
29	41	46.1	192	20 AAY01971	HIV-1 viral infect
30	41	46.1	192	20 AAY01972	HIV-1 viral infect
31	41	46.1	203	14 AAR43876	HTLV-III SOR gene
32	41	46.1	206	21 AAG45698	Arabidopsis thalia
33	41	46.1	208	21 AAG40713	zea mays protein f
34	41	46.1	342	19 AAW98834	H. pylori GHP0 150
35	41	46.1	838	21 AAG35855	Arabidopsis thalia
36	41	46.1	843	21 AAG35854	Arabidopsis thalia
37	41	46.1	849	21 AAG35853	Arabidopsis thalia
38	40	44.9	18	16 AAR78776	Protegrin peptide
39	40	44.9	26	20 AAY18363	Human chromogranin
40	40	44.9	31	20 AAY18362	Human chromogranin
41	40	44.9	112	21 AAB41829	Human ORF1593
42	40	44.9	374	15 AAR62758	TcTA sequence. Sa
43	40	44.9	439	21 AAY53797	Amino acid sequenc
44	39	43.8	114	21 AAG05624	Arabidopsis thalia
45	39	43.8	119	21 AAG05623	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

AAW99404  
ID AAW99404 standard; peptide; 18 AA.

XX AAW99404;

XX AC

XX DT 08-JUN-1999 (first entry)

XX DE Protegrin derivative peptide SM1727.

XX Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

XX OS Synthetic.

XX PN WO9907728-A2.

XX PD 18-FEB-1999.

XX PF 06-AUG-1998; 98WO-FR01757.

XX PR 12-AUG-1997; 97FR-0010297.

XX PA (SYNT-) SYNT:EM SA.

XX PI Calas E, Chavanieu A, Grassy G, Kaczorek M;

XX DR WPI; 1999-190034/16.

XX PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
PT as carriers to deliver active agents into cells

PS Claim 7; Page 28; 37pp; French.  
XX This peptide represents a linear derivative of the protegrin family of  
CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
CC antibiotic family which contain a beta-sheet secondary structure linked  
CC by disulphide bridges. The new derivatives are linear and lack the  
CC disulphide bridge. The novel derivatives are used to deliver active  
CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
CC antivirals and anti-inflammatories, etc. The derivatives are non-toxic  
CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
CC including crossing the blood-brain barrier.  
XX Sequence 18 AA;

Query Match 100.0%; Score 89; DB 20; Length 18;  
Best Local Similarity 100.0%; Pred. No. 3.9e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGVSVSFRRRSYSLRGR 18  
Db 1 rgvsvsfrrrsyrlgr 18

RESULT 2  
AAW36214  
ID AAW36214 standard; peptide; 18 AA.  
XX  
AC AAW36214;  
XX

DT 12-FEB-1998 (first entry)  
XX

XX Antimicrobial protegrin peptide (14).  
DE Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
KW retrovirus; HIV; human immunodeficiency virus; preservation;  
KW disjunction; prophylaxis; treatment; infection; disease;  
KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
KW respiratory infection; urinary tract infection; MRSA; protozoan;  
KW vancomycin resistant enterococcus; pathogen; multi-drug resistance;  
KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
XX Synthetic.  
OS  
OS Sus scrofa.  
XX

FT Key Location/Qualifiers  
FT Misc-difference 1 /note= "D-form residue"  
FT Misc-difference 2 /note= "D-form residue"  
FT Misc-difference 3 /note= "D-form residue"  
FT Misc-difference 4 /note= "D-form residue"  
FT Misc-difference 5 /note= "D-form residue"  
FT Misc-difference 6 /note= "D-form residue"  
FT Misc-difference 7 /note= "D-form residue"  
FT Misc-difference 8 /note= "D-form residue"  
FT Misc-difference 9 /note= "D-form residue"  
FT Misc-difference 10 /note= "D-form residue"  
FT Misc-difference 11 /note= "D-form residue"

FT Misc-difference 12 /note= "D-form residue"  
FT Misc-difference 13 /note= "D-form residue"  
FT Misc-difference 14 /note= "D-form residue"  
FT Misc-difference 15 /note= "D-form residue"  
FT Misc-difference 16 /note= "D-form residue"  
FT Misc-difference 17 /note= "D-form residue"  
FT Misc-difference 18 /note= "D-form residue"  
XX WO9718826-A1.  
XX  
PD 29-MAY-1997.  
XX  
PF 22-NOV-1996; 96WO-US18544.  
XX  
PR 21-NOV-1996; 96US-0752852.  
PR 22-NOV-1995; 95US-0562346.  
PR 17-MAY-1996; 96US-0649811.  
PR 01-AUG-1996; 96US-0690921.  
XX  
PA (INTR-) INTRABIOTICS PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
XX WPI; 1997-297871/27.  
XX  
PT New antimicrobial protegrin peptide(s) - having activity against  
PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
PT (e.g. HIV)  
XX  
PS Claim 23; Page 104; 130pp; English.  
XX  
CC The present sequence is an antimicrobial protegrin peptide, which  
CC has a broad spectrum of activity against microbial targets,  
CC including gram-positive and gram-negative bacteria, yeast, fungi,  
CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
CC It can be used to preserve or disinfect a variety of materials,  
CC including medical equipment, foodstuffs, cosmetics, contact lens  
CC solutions, medicaments or other nutrient containing materials. It  
CC can also be used for the prophylaxis or treatment of microbial  
CC infections or diseases in plants and animals, e.g. conjunctivitis,  
CC keratitis, corneal ulcers, stomach ulcers associated with  
CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
CC sepsis, endocarditis, pneumonia and other respiratory infections,  
CC urinary tract infections, systemic candidiasis and oral mucositis.  
CC It is biostatic or biocidal against clinically relevant pathogens  
CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
CC Enterococcus faecium or faecalis, penicillin resistant  
CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
CC to 1 mg/kg/day, by injection.  
XX  
SQ Sequence 18 AA;

Query Match 77.5%; Score 69; DB 18; Length 18;  
Best Local Similarity 77.8%; Pred. No. 7.4e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 RGVSVSFRRRSYSLRGR 18  
Db 1 rgvsvsfrrrsyrlgr 18  
RESULT 3

AAW36213  
ID AAW36213 standard; peptide; 17 AA.  
XX  
AC AAW36213;  
XX  
DT 12-FEB-1998 (first entry)  
XX  
DE Antimicrobial protegrin peptide (13).  
XX  
KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
KW retrovirus; HIV; human immunodeficiency virus; preservation;  
KW disinfection; prophylaxis; treatment; infection; disease;  
KW Helicobacter pylori; sexually transmitted disease; STD;  
KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
KW respiratory infection; urinary tract infection; MRSA; protozoan;  
KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
XX  
OS Synthetic.  
OS Sus scrofa.  
XX  
PN W09718826-A1.  
XX  
PD 29-MAY-1997.  
XX  
XX 22-NOV-1996; 96WO-US18544.  
XX  
PR 21-NOV-1996; 96US-0752852.  
PR 22-NOV-1995; 95US-0562346.  
PR 17-MAY-1996; 96US-0649811.  
PR 01-AUG-1996; 96US-0690921.  
XX  
PA (INTR-) INTRABIOTICS PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
XX  
XX Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
XX WPI; 1997-297871/27.  
XX  
XX New antimicrobial protegrin peptide(s) - having activity against  
PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
PT (e.g. HIV)  
XX  
PS Claim 23; Page 104; 130pp; English.  
XX  
CC The present sequence is an antimicrobial protegrin peptide, which  
CC has a broad spectrum of activity against microbial targets,  
CC including gram-positive and gram-negative bacteria, yeast, fungi,  
CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
CC It can be used to preserve or disinfect a variety of materials,  
CC including medical equipment, foodstuffs, cosmetics, contact lens  
CC solutions, medicaments or other nutrient containing materials. It  
CC can also be used for the prophylaxis or treatment of microbial  
CC infections or diseases in plants and animals, e.g. conjunctivitis,  
CC keratitis, corneal ulcers, stomach ulcers associated with  
CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
CC sepsis, endocarditis, pneumonia and other respiratory infections,  
CC urinary tract infections, systemic candidiasis and oral mucositis.  
CC It is biostatic or biocidal against clinically relevant pathogens  
CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
CC Enterococcus faecium or faecalis, penicillin resistant  
CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
CC to 1 mg/kg/day, by injection.  
XX  
SQ Sequence 17 AA;

Query Match 61.2%; Score 54.5; DB 18; Length 17;  
Best Local Similarity 72.2%; Pred. No. 0.016;

Matches 13; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1 RGVSVFRRRSYSLRGGR 18  
| | | | | | | | | |  
Db 1 Rgvvcfrrcy-crggr 17  
RESULT 4  
AAW36209  
ID AAW36209 standard; peptide; 15 AA.  
XX  
AC AAW36209;  
XX  
DT 12-FEB-1998 (first entry)  
XX  
DE Antimicrobial protegrin peptide (9).  
XX  
KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
KW retrovirus; HIV; human immunodeficiency virus; preservation;  
KW disinfection; prophylaxis; treatment; infection; disease;  
KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
KW respiratory infection; urinary tract infection; MRSA; protozoan;  
KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
XX  
OS Synthetic.  
OS Sus scrofa.  
XX  
PN W09718826-A1.  
XX  
PD 29-MAY-1997.  
XX  
XX 22-NOV-1996; 96WO-US18544.  
XX  
XX 21-NOV-1996; 96US-0752852.  
PR 22-NOV-1995; 95US-0562346.  
PR 17-MAY-1996; 96US-0649811.  
PR 01-AUG-1996; 96US-0690921.  
XX  
PA (INTR-) INTRABIOTICS PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
XX  
XX Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
XX WPI; 1997-297871/27.  
XX  
XX New antimicrobial protegrin peptide(s) - having activity against  
PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
PT (e.g. HIV)  
XX  
PS Claim 23; Page 104; 130pp; English.  
XX  
CC The present sequence is an antimicrobial protegrin peptide, which  
CC has a broad spectrum of activity against microbial targets,  
CC including gram-positive and gram-negative bacteria, yeast, fungi,  
CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
CC It can be used to preserve or disinfect a variety of materials,  
CC including medical equipment, foodstuffs, cosmetics, contact lens  
CC solutions, medicaments or other nutrient containing materials. It  
CC can also be used for the prophylaxis or treatment of microbial  
CC infections or diseases in plants and animals, e.g. conjunctivitis,  
CC keratitis, corneal ulcers, stomach ulcers associated with  
CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
CC sepsis, endocarditis, pneumonia and other respiratory infections,  
CC urinary tract infections, systemic candidiasis and oral mucositis.  
CC It is biostatic or biocidal against clinically relevant pathogens  
CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
CC Enterococcus faecium or faecalis, penicillin resistant  
CC Streptococcus pneumoniae and methicillin resistant Staphylococcus

CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX Sequence 15 AA;  
 SQ

Query Match 52.8%; Score 47; DB 18; Length 15;  
 Best Local Similarity 71.4%; Pred. No. 0.24;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYSL 14  
 ||| | |||| |  
 Db 1 rgvcvcfrrrcycl 14

RESULT 5  
 AAW36210  
 ID AAW36210 standard; peptide; 15 AA.

XX AC AAW36210;

DT 12-FEB-1998 (first entry)

DE Antimicrobial protegrin peptide (10).

XX Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.

OS Synthetic.  
 OS Sus scrofa.

PN WO9718826-A1.

PD 29-MAY-1997.

PF 22-NOV-1996; 96WO-US18544.

PR 21-NOV-1996; 96US-0752852.

PR 22-NOV-1995; 95US-0562346.

PR 17-MAY-1996; 96US-0649811.

PR 01-AUG-1996; 96US-0690921.

XX (INTR-) INTRABIOTICS PHARM INC.

PA (REGC ) UNIV CALIFORNIA.

XX Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
 PI WPI; 1997-297871/27.

DR New antimicrobial protegrin peptide(s) - having activity against

PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)

PS Claim 23; Page 104; 130pp; English.

XX The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,

CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.

XX Sequence 15 AA;

Query Match 52.8%; Score 47; DB 18; Length 15;  
 Best Local Similarity 71.4%; Pred. No. 0.24;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGVSVSFRRRSYSL 14  
 ||| | |||| |  
 Db 1 rgvcvcfrrrcycl 14

RESULT 6

AAW36215  
 ID AAW36215 standard; peptide; 15 AA.

XX AC AAW36215;

DT 12-FEB-1998 (first entry)

DE Antimicrobial protegrin peptide (15).

XX Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.

OS Synthetic.

OS Sus scrofa.

PN WO9718826-A1.

PD 29-MAY-1997.

PF 22-NOV-1996; 96WO-US18544.

PR 21-NOV-1996; 96US-0752852.

PR 22-NOV-1995; 95US-0562346.

PR 17-MAY-1996; 96US-0649811.

PR 01-AUG-1996; 96US-0690921.

XX (INTR-) INTRABIOTICS PHARM INC.

PA (REGC ) UNIV CALIFORNIA.

XX Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
 PI WPI; 1997-297871/27.

DR New antimicrobial protegrin peptide(s) - having activity against

PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)

PS Claim 23; Page 104; 130pp; English.

XX The present sequence is an antimicrobial protegrin peptide, which

CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC *Helicobacter pylori*, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections,  
 CC urinary tract infections, systemic candidiasis and oral mucositis.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC *Enterococcus faecium* or faecalis, penicillin resistant  
 CC *Streptococcus pneumoniae* and methicillin resistant *Staphylococcus*  
 CC *aureus* (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
 CC to 1 mg/kg/day, by injection.  
 XX  
 SQ Sequence 15 AA;

Query Match 52.8%; Score 47; DB 18; Length 15;  
 Best Local Similarity 71.4%; Pred. No. 0.24; Mismatches 4; Indels 0; Gaps 0;  
 Matches 10; Conservative 0;

QY 1 RGVSVSFRRSYSL 14  
 ||| | ||| | |  
 Db 1 rgvcvfcrrcyl 14

## RESULT 7

AA77962  
 ID AAY77962 standard; protein; 366 AA.

AC AAY77962;

DT 14-JUN-2000 (first entry)

DE A. thaliana environmental stress tolerance related protein.

KW Environmental stress; plant; transgenic plant; anaerobic; flooding; cold;  
 KW dehydration; drought; heat stress; salinity; osmotolerance.

OS Arabidopsis thaliana.

PN WO200008187-A2.

PD 17-FEB-2000.

PF 04-AUG-1999; 99WO-EP05652.

PR 04-AUG-1998; 98EP-0202634.

PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

PI Lee JH, Verbruggen N;

DR WPI: 2000-205726/18.  
 DR N-PSDB; AA298342.

PT Isolating of polynucleic acids useful for producing transgenic plant by  
 PT isolating genes involved in tolerance to environmental stress

PS Claim 12; Page 236-238; 312pp; English.

XX The invention relates to isolation of coding sequences and/or genes  
 CC involved in tolerance to environmental stress in plants. The sequences  
 CC (AA298305-298365) are useful for producing a transgenic plant having  
 CC enhanced tolerance or resistance to environmental stress conditions such  
 CC as anaerobic, flooding, cold, dehydration, drought, heat stress or  
 CC salinity. This is useful for producing improved yield, growth,  
 CC development and productivity under environmental stress conditions, and

CC also provides growth of crops in areas where they cannot grow without  
 CC the induced osmotolerance. Sequences AAY77925-984 represent polypeptide  
 CC sequences from A. thaliana that are encoded by the genes involved in  
 CC environmental stress tolerance.  
 XX  
 SQ Sequence 366 AA;

Query Match 47.2%; Score 42; DB 21; Length 366;  
 Best Local Similarity 60.0%; Pred. No. 48;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 SVSFRRSYSLRGR 18  
 ||:|||| | | |  
 Db 52 svnyrrrelslgr 66

## RESULT 8

AAB86197  
 ID AAB86197 standard; Protein; 3025 AA.

XX AAB86197;

DT 21-AUG-2001 (first entry)

DE HIV-1 subtype C protein fragment #3.

KW Infection; diagnosis; human; humoral immune response; antiviral;  
 KW cellular immune response; vaccine; treatment; gene therapy.

OS Human immunodeficiency virus type 1.

FH Key Location/Qualifiers

FT Misc-difference 1..3025  
 FT /label= xaa  
 FT /note= "Xaa represents a stop codon"

PN DE10056747-A1.

PD 31-MAY-2001.

PF 16-NOV-2000; 2000DE-1056747.

PR 16-NOV-1999; 99DE-1055089.

XX (SHAO/) SHAO Y.

PA (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.

PI Wagner R, Wolf H, Shao Y, Graf M;

XX WPI: 2001-336417/36.

PT New nucleic acid sequences from a human immune deficiency virus  
 PT intersubtype, useful for treatment, prevention and diagnosis of  
 PT infection

PS Disclosure; Fig 8A-O; 48pp; German.

XX This invention describes a novel polynucleotide isolated from human  
 CC immunodeficiency virus type 1 subtype C/B' which can be used for the  
 CC induction of specific humoral and cellular immune responses. (I) and  
 CC polypeptides (II) encoded by them, are useful in pharmaceuticals,  
 CC vaccines and diagnostic agents, particularly for treatment or prevention  
 CC of human immune deficiency virus-1 (HIV-1) infections, also for rational  
 CC design of test or therapeutic reagents, or gene therapy vectors.  
 CC Polypeptides, especially antibodies, specifically directed against (II)  
 CC are similarly useful as pharmaceutical and diagnostic agents. (I) are  
 CC specific for intersubtype C/B' of HIV-1 so are useful in regions  
 CC (particularly China and South-East Asia) where this subtype is prevalent.  
 CC The products of the invention have antiviral activity. This sequence  
 CC represents a protein encoded by the HIV-1 subtype C genome described in  
 CC the method of the invention.

```

SQ      Sequence      3025 AA;

Query Match      47.2%; Score 42; DB 22; Length 3025;
Best Local Similarity 46.7%; Pred. NO. 4.6e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      1  RGVSVSFRRSYSLR 15
Db      557  KGSIRYRSYSLR 571

RESULT      9
AAG25990
ID      AAG25990 standard; Protein; 59 AA.
XX
AC      AAG25990;
XX
DT      17-OCT-2000 (first entry)
XX
DE      Zea mays protein fragment SEQ ID NO: 30276.
XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence; corn.
XX
OS      Zea mays subsp. mays.
XX
PN      EP1033405-A2.
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      30-APR-1999; 99US-0132407.
PR      04-MAY-1999; 99US-0132484.
PR      05-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      06-MAY-1999; 99US-0132487.
PR      07-MAY-1999; 99US-0132863.
PR      11-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      14-MAY-1999; 99US-0134370.
PR      18-MAY-1999; 99US-0134768.
PR      19-MAY-1999; 99US-0134941.
PR      20-MAY-1999; 99US-0135124.
PR      21-MAY-1999; 99US-0135353.
PR      24-MAY-1999; 99US-0135629.
PR      25-MAY-1999; 99US-0136021.
PR      27-MAY-1999; 99US-0136392.
PR      28-MAY-1999; 99US-0136782.
PR      01-JUN-1999; 99US-0137222.
PR      03-JUN-1999; 99US-0137528.
PR      04-JUN-1999; 99US-0137502.
PR      07-JUN-1999; 99US-0137724.
PR      08-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      10-JUN-1999; 99US-0138847.
PR      14-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      16-JUN-1999; 99US-0139453.
PR      17-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139763.
PR      21-JUN-1999; 99US-0139817.
PR      22-JUN-1999; 99US-0139899.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.
PR      01-JUL-1999; 99US-0141842.
PR      01-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144085.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144325.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      19-JUL-1999; 99US-0144335.
PR      20-JUL-1999; 99US-0144352.
PR      20-JUL-1999; 99US-0144632.
PR      20-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145088.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145218.
PR      23-JUL-1999; 99US-0145224.
PR      26-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
PR      02-AUG-1999; 99US-0146389.
PR      03-AUG-1999; 99US-0147038.
PR      04-AUG-1999; 99US-0147204.
PR      04-AUG-1999; 99US-0147302.
PR      05-AUG-1999; 99US-0147192.
PR      05-AUG-1999; 99US-0147260.
PR      06-AUG-1999; 99US-0147303.
PR      06-AUG-1999; 99US-0147416.

```

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 15-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151438.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 46.1%; Score 41; DB 21; Length 59;  
Best Local Similarity 53.3%; Pred. No. 10;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Caps 0;

QY 2 GVSVSFRRRSYSLRG 16  
|:|:| |:|:| |:|  
Db 28 gisvdrhrknksleg 42

RESULT 10  
AAG45699  
ID AAG45699 standard; Protein; 141 AA.  
XX AAG45699;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 57406.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 46.1%; Score 41; DB 21; Length 141;  
Best Local Similarity 53.3%; Pred. No. 25;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GVSVSFRRRSYSLRG 16  
|:|:| |:|:| |:  
Db 28 gisvdhrrkrsleg 42

RESULT 11  
AAG40714



ID	AAG40714 standard; Protein; 142 AA.		
XX		PR	18-JUN-1999;
AC	AAG40714;	PR	18-JUN-1999;
XX		PR	99US-0139462.
DT		PR	99US-0139463.
XX	18-OCT-2000 (first entry)	PR	99US-0139750.
XX		PR	99US-0139763.
DE	Zea mays protein fragment SEQ ID NO: 50555.	PR	99US-0139817.
XX		PR	99US-0139859.
KW	Protein identification; signal transduction pathway; metabolic pathway;	PR	99US-0140354.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	PR	99US-0140353.
KW	termination sequence; corn.	PR	99US-0140695.
XX		PR	99US-0140823.
OS		PR	99US-0140991.
XX	Zea mays subsp. mays.	PR	99US-0141287.
PN	EP1033405-A2.	PR	99US-0141842.
XX		PR	99US-0142154.
XX		PR	99US-0142055.
PD	06-SEP-2000.	PR	99US-0142390.
XX		PR	99US-0142803.
PF	25-FEB-2000; 2000EP-0301439.	PR	99US-0142920.
XX		PR	99US-0142977.
PR	25-FEB-1999;	PR	99US-0143542.
PR	99US-0121825.	PR	99US-0143542.
PR	05-MAR-1999;	PR	99US-0143624.
PR	99US-0123180.	PR	99US-0144085.
PR	09-MAR-1999;	PR	99US-0144085.
PR	99US-0123548.	PR	99US-0144325.
PR	23-MAR-1999;	PR	99US-0144325.
PR	99US-0125788.	PR	99US-0144331.
PR	25-MAR-1999;	PR	99US-0144332.
PR	29-MAR-1999;	PR	99US-0144333.
PR	99US-0126264.	PR	99US-0144334.
PR	01-APR-1999;	PR	99US-0144335.
PR	99US-0126785.	PR	99US-0144352.
PR	08-APR-1999;	PR	99US-0144632.
PR	99US-0128234.	PR	99US-0144884.
PR	16-APR-1999;	PR	99US-0144884.
PR	99US-0128714.	PR	99US-0145086.
PR	21-APR-1999;	PR	99US-0145086.
PR	99US-0129845.	PR	99US-0145088.
PR	23-APR-1999;	PR	99US-0145085.
PR	99US-0130077.	PR	99US-0145087.
PR	25-APR-1999;	PR	99US-0145089.
PR	99US-0130510.	PR	99US-0145089.
PR	28-APR-1999;	PR	99US-0145192.
PR	99US-0131449.	PR	99US-0145145.
PR	30-APR-1999;	PR	99US-0145218.
PR	99US-0132048.	PR	99US-0145224.
PR	03-MAY-1999;	PR	99US-0145224.
PR	99US-0132407.	PR	99US-0145276.
PR	04-MAY-1999;	PR	99US-0145913.
PR	99US-0132484.	PR	99US-0145918.
PR	05-MAY-1999;	PR	99US-0145919.
PR	99US-0132485.	PR	99US-0145951.
PR	06-MAY-1999;	PR	99US-0146386.
PR	99US-0132486.	PR	99US-0146388.
PR	07-MAY-1999;	PR	99US-0146389.
PR	99US-0132863.	PR	99US-0147038.
PR	11-MAY-1999;	PR	99US-0147204.
PR	99US-0134256.	PR	99US-0147204.
PR	14-MAY-1999;	PR	99US-0147302.
PR	99US-0134218.	PR	99US-0147192.
PR	14-MAY-1999;	PR	99US-0147260.
PR	99US-0134221.	PR	99US-0147303.
PR	14-MAY-1999;	PR	99US-0147303.
PR	99US-0134370.	PR	99US-0147416.
PR	18-MAY-1999;	PR	99US-0147416.
PR	99US-0134768.	PR	99US-0147935.
PR	19-MAY-1999;	PR	99US-0147935.
PR	99US-0134941.	PR	99US-0148171.
PR	20-MAY-1999;	PR	99US-0148171.
PR	99US-0135124.	PR	99US-0148341.
PR	21-MAY-1999;	PR	99US-0148341.
PR	99US-0135353.	PR	99US-0148565.
PR	24-MAY-1999;	PR	99US-0148565.
PR	99US-0135629.	PR	99US-0149368.
PR	25-MAY-1999;	PR	99US-0149175.
PR	99US-0136021.	PR	99US-0149175.
PR	27-MAY-1999;	PR	99US-0149426.
PR	99US-0136592.	PR	99US-0149426.
PR	28-MAY-1999;	PR	99US-0149722.
PR	99US-0136782.	PR	99US-0149722.
PR	01-JUN-1999;	PR	99US-0149929.
PR	99US-0137222.	PR	99US-0149929.
PR	03-JUN-1999;	PR	99US-0149929.
PR	99US-0137528.	PR	99US-0149929.
PR	04-JUN-1999;	PR	99US-0149929.
PR	99US-0137502.	PR	99US-0149929.
PR	07-JUN-1999;	PR	99US-0149929.
PR	99US-0137724.	PR	99US-0149929.
PR	08-JUN-1999;	PR	99US-0149929.
PR	99US-0138094.	PR	99US-0149929.
PR	10-JUN-1999;	PR	99US-0149929.
PR	99US-0138540.	PR	99US-0149929.
PR	10-JUN-1999;	PR	99US-0149929.
PR	99US-0138847.	PR	99US-0149929.
PR	14-JUN-1999;	PR	99US-0149929.
PR	99US-0139119.	PR	99US-0149929.
PR	16-JUN-1999;	PR	99US-0149929.
PR	99US-0139452.	PR	99US-0149929.
PR	16-JUN-1999;	PR	99US-0149929.
PR	99US-0139453.	PR	99US-0149929.
PR	17-JUN-1999;	PR	99US-0149929.
PR	99US-0139492.	PR	99US-0149929.
PR	18-JUN-1999;	PR	99US-0149929.
PR	99US-0139454.	PR	99US-0149929.
PR	18-JUN-1999;	PR	99US-0149929.
PR	99US-0139455.	PR	99US-0149929.
PR	18-JUN-1999;	PR	99US-0149929.
PR	99US-0139456.	PR	99US-0149929.
PR	18-JUN-1999;	PR	99US-0149929.
PR	99US-0139457.	PR	99US-0149929.
PR	18-JUN-1999;	PR	99US-0149929.
PR	99US-0139458.	PR	99US-0149929.
PR	18-JUN-1999;	PR	99US-0149929.
PR	99US-0139459.	PR	99US-0149929.
PR	18-JUN-1999;	PR	99US-0149929.
PR	99US-0139460.	PR	99US-0149929.



PR 26-SEP-1997; 97US-0060172.  
 XX 18-SEP-1997; 97US-0059283.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 XX Ayyavoo V, Nagashunmugam T, Weiner DB;  
 XX WPI; 1999-263380/22.  
 DR New attenuated vif (viral infectivity factor) genes, used in genetic  
 PT vaccines against HIV-1  
 XX Disclosure; Page 57-58; 92pp; English.  
 PS  
 XX The present sequence represents the consensus sequence derived from  
 CC attenuated, non-functional forms of the accessory protein vif (viral  
 CC infectivity factor) of Human Immunodeficiency virus type 1 (HIV-1).  
 CC The specification describes novel vif proteins (AA01969-88) and the  
 CC genes encoding them (AA035217-36). The vif gene is an accessory gene  
 CC for HIV-1 that has low functional mutagenicity and is conserved. In  
 CC addition, attenuated, non functional vif clones are able to induce  
 CC immune responses capable of destroying native pathogen. Vif nucleic  
 CC acids may be used to immunize mammals. The attenuated, non-functional  
 CC vif genes may be used in concert with other HIV-1 genes to produce  
 CC vaccine that has a broad immune response against all viral  
 CC components, and which mimics many aspects of the immune responses  
 CC induced by a live attenuated virus. Prophylactic vaccines which include  
 CC vif could limit both viral escape and contribute to lowering the viral  
 CC set point during early infection stages.  
 XX  
 SQ Sequence 190 AA;

Query Match 46.1%; Score 41; DB 20; Length 190;  
 Best Local Similarity 53.8%; Pred. No. 35;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRSYS 13  
 :|||: :|:| |  
 Db 81 qgvslwkrkrrys 93

RESULT 14  
 AAP81855  
 ID AAP81855 standard; protein; 192 AA.  
 XX  
 AC AAP81855;  
 XX  
 DT 16-DEC-1990 (first entry)  
 XX  
 DE Sequence encoded by LAV EL I Q gene.  
 XX  
 KW HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.  
 XX  
 OS Lymphadenopathy associated virus EL I.  
 XX  
 PN W08707906-A.  
 XX  
 PD 30-DEC-1987.  
 XX  
 PF 22-JUN-1987; 87WO-EF00326.  
 XX  
 PR 23-JUN-1986; 86EP-0401380.  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA  
 XX Allison M, Sonigo P, Wain-Hobson S, Montagnier L;  
 XX WPI; 1988-014396/02.  
 DR N-PSDB; AAN80436.  
 XX  
 XX New variants of lymphadenopathy associated virus (LAV) -  
 PT used for prodn. of DNA, antigens and antibodies used in

PT diagnosis of AIDS and pre-AIDS  
 XX  
 PS Claim 8; Fig 7A-7J; 72pp; English.  
 XX  
 CC LAV EL I (AAN80436) and LAV MA L (AAN80437) were isolated from the  
 CC peripheral blood lymphocytes of patients. Different AIDS virus isolates  
 CC concerned are designated by 3 letters of the patients name. Stable probes  
 CC including the DNA sequences can be used for detection of the new LAV  
 CC viruses or related viruses or DNA proviruses in eg. biological samples.  
 CC The proteins or peptides can be used for detection of antibodies induced  
 CC in vivo and present in biological fluids. The DNA can also be used for  
 CC the expression of LAV viral antigens for the prodn. of a vaccine against  
 CC LAV. The polypeptides can also be used for the prodn. of antibodies for  
 CC the detection of proteins related to the LAV viruses, partic. for  
 CC diagnosis of AIDS or pre-AIDS.  
 XX  
 SQ Sequence 192 AA;

Query Match 46.1%; Score 41; DB 9; Length 192;  
 Best Local Similarity 53.8%; Pred. No. 35;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSFRRSYS 13  
 :|||: :|:| |  
 Db 83 qgvslwkrkrrys 95

RESULT 15  
 AAR09302  
 ID AAR09302 standard; protein; 192 AA.  
 XX  
 AC AAR09302;  
 XX  
 DT 27-FEB-1991 (first entry)  
 XX  
 DE Sequence deduced from vif gene of HIV 1-NDK.  
 XX  
 KW Human Immunodeficiency virus; AIDS.  
 XX  
 OS HIV 1-NDK.  
 XX  
 PN W09013630-A.  
 XX  
 PD 15-NOV-1990.  
 XX  
 PF 02-MAY-1990; 90WO-FR00312.  
 XX  
 PR 03-MAY-1989; 89FR-0005914.  
 XX  
 XX (INRM ) INSERM INST NAT SANTE.  
 PA  
 XX Barre-Sinoussi F, Chermann JC, Devaux C, Rey F, Sire J;  
 PI Spire B;  
 XX  
 DR WPI; 1990-361470/48.  
 DR N-PSDB; AAO06635.  
 XX  
 XX New HIV-NDK retrovirus and protein component - used in vaccines  
 PT against immuno-deficiency disorders and in raising MABS for  
 PT retro-virus detection in vivo.  
 XX  
 PS Disclosure; Fig 2; 37pp; French.  
 XX  
 CC The HIV NDK virus was isolated from peripheral blood lymphocytes of  
 CC an AIDS patient. A genomic library was prepd. from DNA extracted  
 CC from CEM cells infected with the virus. The library was screened  
 CC with a pBT1 probe corresp. to a fragment from HIV 1. The virus is  
 CC more cytopathic than other strains and is not inhibited by OKT4A.  
 CC It has been deposited as CNCM I-857. The sequence can be used to  
 CC express proteins useful for diagnosing the presence of NDK and  
 CC related viruses and in vaccines against immunodeficiency diseases.  
 CC See also AAR09301-4.

xx  
SQ Sequence 192 AA;

Query Match 46.1%; Score 41; DB 11; Length 192;  
Best Local Similarity 53.8%; Pred. No. 35;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGVSVSPRRSYS 13  
Db :|||: ||| ||  
83 qgvsiewkrkrys 95

Search completed: February 12, 2002, 12:30:29  
Job time: 362 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:26:07 ; Search time 232.64 Seconds  
(without alignments)  
11.317 Million cell updates/sec

Title: US-09-485-571-15

Perfect score: 89

Sequence: 1 RGRRLSYRRRFSVSVGR 18.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	58.4	101	12	Q9ICW8
2	52	58.4	101	12	Q9ICW2
3	51	57.3	92	12	Q38024
4	49	55.1	93	12	Q73508
5	49	55.1	93	12	Q73509
6	49	55.1	93	12	Q73512
7	49	55.1	93	12	Q73514
8	49	55.1	93	12	Q73527
9	49	55.1	93	12	Q73529
10	49	55.1	93	12	Q73531
11	49	55.1	93	12	Q93139
12	49	55.1	94	12	Q73525
13	49	55.1	94	12	Q86541
14	48	53.9	94	12	O41486
15	48	53.9	108	12	Q9QNN0
16	45	50.6	61	6	Q9GLQ9
17	45	50.6	61	6	Q9GLQ3
18	45	50.6	61	6	Q9GLP8
19	45	50.6	61	6	Q9GJQ1

20 45 50.6 62 6 Q9TUC2  
21 45 50.6 62 6 Q9GLQ0  
22 45 50.6 63 6 Q9TUC4  
23 45 50.6 65 6 Q9GLQ2  
24 45 50.6 104 12 Q9E9R1  
25 45 50.6 173 2 Q9ETAR8  
26 44 49.4 140 12 Q36986  
27 44 49.4 627 10 Q9LPN2  
28 44 49.4 685 10 Q9SS80  
29 43 48.3 117 11 Q9D596  
30 43 48.3 137 13 Q9PSV7  
31 43 48.3 1016 5 O17484  
32 43 48.3 1016 5 O17485  
33 42 47.2 94 12 O41484  
34 42 47.2 118 1 O9YEK7  
35 42 47.2 317 8 O79545  
36 42 47.2 326 10 Q9AUN0  
37 42 47.2 1138 10 Q9ZW00  
38 42 47.2 2086 11 Q9QXP5  
39 42 47.2 2404 11 Q9QX47  
40 41 46.1 186 2 Q52604  
41 41 46.1 186 2 Q9R430  
42 41 46.1 374 11 Q9JJL7  
43 41 46.1 654 2 Q9ANS5  
44 41 46.1 1696 4 Q9Y4F4  
45 40 44.9 105 12 Q9QBT8

Q9TUC2 sminthopsis  
Q9GLQ0 bettongia p  
Q9TUC4 sminthopsis  
Q9GLQ2 lagorcheste  
Q9E9R1 hop latent  
Q9ETAR8 corynebacte  
Q36986 lily sympto  
Q9LPN2 arabidopsis  
Q9SS80 arabidopsis  
Q9D596 mus musculu  
Q9PSV7 conger myrl  
O17484 plodia inte  
O17485 plodia inte  
O41484 potato viru  
Q9YEK7 aeropyrum p  
O79545 scenedemus  
Q9AUN0 cryza sativ  
Q9ZW00 arabidopsis  
Q9QXP5 mus musculu  
Q9QX47 mus musculu  
Q52604 agrobacteri  
Q9R430 agrobacteri  
Q9JJL7 rattus norv  
Q9ANS5 rhodobacter  
Q9Y4F4 homo sapien  
Q9QBT8 potato roug

#### ALIGNMENTS

RESULT 1  
Q9ICW8 ID Q9ICW8 PRELIMINARY; PRT; 101 AA.  
AC Q9ICW8  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL 11.6 KDA PROTEIN (FRAGMENT).  
OS Kalanchoe latent virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=132477;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1452;  
RA Nicolaissen M., Nielsen S.L.;  
RT "Nucleotide sequence of the 3'- terminal region of Kalanchoe latent virus."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ293570; CAB97501.2; -.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
KW Hypothetical protein.  
FT NON\_TER 101  
SQ SEQUENCE 101 AA; 13556 MW; C55F15AA3718BEDD CRC64;

Query Match 58.4%; Score 52; DB 12; Length 101;  
Best Local Similarity 56.2%; Pred. No. 0.3;  
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 3 GRLSYRRRFSVSVGR 18  
||:||||:|:|:  
DB 42 GRSTYARRRAISIGR 57

RESULT 2  
Q9ICW2 ID Q9ICW2 PRELIMINARY; PRT; 101 AA.  
AC Q9ICW2  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE HYPOTHETICAL 11.6 KDA PROTEIN (FRAGMENT).  
OS Kalanchoe latent virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=132477;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PV-0290;  
RA Nicolaïsen M., Nielsen S.L.;  
RT "Nucleotide sequence of the 3'- terminal region of Kalanchoe latent virus".  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ293571; CAB97507.2; -.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
KW Hypothetical protein.  
FT NON\_TER 101 101  
SQ SEQUENCE 101 AA; 11591 MW; 39BD224E73E3DBFA CRC64;

Query Match 58.4%; Score 52; DB 12; Length 101;  
Best Local Similarity 56.2%; Pred. No. 0.3;  
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 3 GRLSYRRRFSVSGR 18  
|||:||||:|:|  
DB 42 GRSTYARRRAISGR 57

RESULT 3  
Q38024 PRELIMINARY; PRT; 92 AA.  
ID O38024  
AC O38024;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE HYPOTHETICAL 10.3 KDA PROTEIN.  
OS Potato virus M.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=12167;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IDAHO;  
RA Cavilleir T.D., Corsini D.L., Berger P.H.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF023877; AAB81273.1; -.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 92 AA; 10319 MW; B0A9AC70B579A980 CRC64;

Query Match 57.3%; Score 51; DB 12; Length 92;  
Best Local Similarity 52.9%; Pred. No. 0.39;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GRLSYRRRFSVSGR 18  
|||:||||:|:|  
DB 40 GGRSKYARRRAAAGR 56

RESULT 4  
O73508 PRELIMINARY; PRT; 93 AA.  
ID O73508  
AC O73508;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE 11K PROTEIN (FRAGMENT).  
OS Potato virus S.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=12169;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=ASCHERSLEBEN;  
RA Matousek J., Schubert J., Dedic P.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y15613; CAA75702.1; -.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
FT NON\_TER 93 93  
SQ SEQUENCE 93 AA; 10521 MW; 2DA6E2925A8C475D CRC64;

Query Match 55.1%; Score 49; DB 12; Length 93;  
Best Local Similarity 52.9%; Pred. No. 0.86;  
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 GRLSYRRRFSVSGR 18  
|||:||||:|:|  
DB 38 GGRSTYARRRARSIGR 54

RESULT 5  
O73509 PRELIMINARY; PRT; 93 AA.  
ID O73509  
AC O73509;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE 11K PROTEIN (FRAGMENT).  
OS Potato virus S.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=12169;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ASCHERSLEBEN;  
RA Matousek J., Schubert J., Dedic P.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y15615; CAA75706.1; -.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
FT NON\_TER 93 93  
SQ SEQUENCE 93 AA; 10521 MW; 2DA6E2925A8C475D CRC64;

Query Match 55.1%; Score 49; DB 12; Length 93;  
Best Local Similarity 52.9%; Pred. No. 0.86;  
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 GRLSYRRRFSVSGR 18  
|||:||||:|:|  
DB 38 GGRSTYARRRARSIGR 54

RESULT 6  
O73512 PRELIMINARY; PRT; 93 AA.  
ID O73512  
AC O73512;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE 11K PROTEIN (FRAGMENT).  
OS Potato virus S.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=12169;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KARLA;  
RA Matousek J., Schubert J., Dedic P.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y15611; CAA75698.1; -.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
FT NON\_TER 93 93  
SQ SEQUENCE 93 AA; 10537 MW; AC2FE2A0F98659B9 CRC64;

Query Match 55.1%; Score 49; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 0.86;  
 Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 GGRLSYRRRRFSVSGR 18  
 ||| :|:|:| :|:|  
 Db 38 GGRSTYARRRRARSIGR 54

RESULT 7  
 OY3514  
 ID O73514 PRELIMINARY; PRT; 93 AA.  
 AC O73514;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KARLA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15612; CAA75700.1; -.  
 DR InterPro: IPR002568; Carla\_C4.  
 DR Pfam: PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10538 MW; 11BD9C8C9957BB85 CRC64;

Query Match 55.1%; Score 49; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 0.86;  
 Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 GGRLSYRRRRFSVSGR 18  
 ||| :|:|:| :|:|  
 Db 38 GGRSTYARRRRARSIGR 54

RESULT 8  
 OY3527  
 ID O73527 PRELIMINARY; PRT; 93 AA.  
 AC O73527;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15609; CAA75694.1; -.  
 DR InterPro: IPR002568; Carla\_C4.  
 DR Pfam: PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10333 MW; AB44F80C8018A5C9 CRC64;

Query Match 55.1%; Score 49; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 0.86;  
 Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 GGRLSYRRRRFSVSGR 18  
 ||| :|:|:| :|:|  
 Db 38 GGRSTYARRRRARSIGR 54

RESULT 9  
 OY3529  
 ID O73529 PRELIMINARY; PRT; 93 AA.  
 AC O73529;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15616; CAA75708.1; -.  
 DR InterPro: IPR002568; Carla\_C4.  
 DR Pfam: PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10507 MW; AC2FE2A0F98645C9 CRC64;

OY 2 GGRLSYRRRRFSVSGR 18  
 ||| :|:|:| :|:|  
 Db 38 GGRSTYARRRRARSIGR 54

Query Match 55.1%; Score 49; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 0.86;  
 Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

RESULT 10  
 OY3531  
 ID O73531 PRELIMINARY; PRT; 93 AA.  
 AC O73531;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).  
 OS Potato virus S.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VITAVA;  
 RA Matousek J., Schubert J., Dedic P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y15610; CAA75696.1; -.  
 DR InterPro: IPR002568; Carla\_C4.  
 DR Pfam: PF01623; Carla\_C4; 1.  
 FT NON\_TER 93  
 SQ SEQUENCE 93 AA; 10333 MW; AB44F80C8018A5C9 CRC64;

Query Match 55.1%; Score 49; DB 12; Length 93;  
 Best Local Similarity 52.9%; Pred. No. 0.86;  
 Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

RESULT 11  
 OY3139  
 ID O93139 PRELIMINARY; PRT; 93 AA.  
 AC O93139;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE 11K PROTEIN (FRAGMENT).

```
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlaviruses.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOBRA;
RA Matousek J., Schubert J., Dedic P.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15614; CAA75704.1; -.
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
FT NON_TER 93
FT SEQUENCE 93 AA; 10565 MW; 11BD8D5D69865A75 CRC64;

Query Match 55.1%; Score 49; DB 12; Length 93;
Best Local Similarity 52.9%; Pred. No. 0.87;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRFSVSGR 18
   ||| :|:| :|:|
Db 38 GGRSTYARKRRARSIGR 54

RESULT 12
ID 073525 PRELIMINARY; PRT; 94 AA.
AC 073525;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 11K PROTEIN.
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlaviruses.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOBRA;
RA Matousek J., Schubert J., Dedic P.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15625; CAA75721.1; -.
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
DR SEQUENCE 94 AA; 10665 MW; 5236BDFD583C830A CRC64;

Query Match 55.1%; Score 49; DB 12; Length 94;
Best Local Similarity 52.9%; Pred. No. 0.87;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRFSVSGR 18
   ||| :|:| :|:|
Db 38 GGRSTYARKRRARSIGR 54

RESULT 13
ID 086541 PRELIMINARY; PRT; 94 AA.
AC 086541;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 11 KDA PROTEIN.
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlaviruses.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033173; PubMed=1413539;
RA Foster G.D., Mills P.R.;
RT "The 3'-nucleotide sequence of an ordinary strain of potato virus S.";
RL Virus Genes 6:213-220(1992).
DR EMBL; S45593; AAB23462.1; -.

```

```
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
DR SEQUENCE 94 AA; 10680 MW; 2241BD8D5D69865A CRC64;

Query Match 55.1%; Score 49; DB 12; Length 94;
Best Local Similarity 52.9%; Pred. No. 0.87;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRFSVSGR 18
   ||| :|:| :|:|
Db 38 GGRSTYARKRRARSIGR 54

RESULT 14
ID 041486 PRELIMINARY; PRT; 94 AA.
AC 041486;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 11K PROTEIN.
OS Potato virus S.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlaviruses.
OX NCBI_TaxID=12169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-RB;
RA Joung Y.H.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U74376; AAB65087.1; -.
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
DR SEQUENCE 94 AA; 10649 MW; C8CCDEF10F00A10A CRC64;

Query Match 53.9%; Score 48; DB 12; Length 94;
Best Local Similarity 52.9%; Pred. No. 1.3;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRFSVSGR 18
   ||| :|:| :|:|
Db 38 GGRSTYGRKRRARSIGR 54

RESULT 15
ID 09QNN0 PRELIMINARY; PRT; 108 AA.
AC 09QNN0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 13K PROTEIN.
OS Potato virus M.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlaviruses.
OX NCBI_TaxID=12167;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOMATO;
RA Grieco F., Di Franco A., Gallitelli D.;
RT "Potato virus M in tomato crops in Southern Italy.";
RL J. Plant Pathol. 78:45-49(1997).
DR EMBL; X85114; CAA59434.1; -.
DR InterPro; IPR002568; Carla_C4.
DR Pfam; PF01623; Carla_C4; 1.
DR SEQUENCE 108 AA; 12182 MW; F84956324A930699 CRC64;

Query Match 53.9%; Score 48; DB 12; Length 108;
Best Local Similarity 52.9%; Pred. No. 1.5;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFRFSVSGR 18

```



Db     |||    |:|||    :|: |  
      40 GGRKYARRRRVISIAR 56

Search completed: February 12, 2002, 12:38:34  
Job time: 747 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:26:27 ; Search time 67.2 Seconds  
(without alignments)  
9.821 Million cell updates/sec

Title: US-09-485-571-15

Perfect score: 89

Sequence: 1 RGGRLSYRRRFSVSGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	69	77.5	149	1 PGI_PIG	P32194 sus scrofa
2	63	70.8	147	1 PGI_PIG	P32195 sus scrofa
3	62	69.7	149	1 PGI_PIG	P32196 sus scrofa
4	62	69.7	149	1 PGI_PIG	P49934 sus scrofa
5	51	57.3	108	1 VNB_PVMG	Q01687 potato viru
6	51	57.3	108	1 VNB_PVMR	P17530 potato viru
7	47	52.8	107	1 VNB_CVB	P37992 chrysanthem
8	45	50.6	59	1 HSP1_MACRU	P42142 macropus ru
9	45	50.6	60	1 HSP1_CAERU	P42131 caenolestes
10	45	50.6	60	1 HSP1_DASVI	P42135 dasyurus vi
11	45	50.6	60	1 HSP1_MACAG	P42137 macropus ag
12	45	50.6	61	1 HSP1_ANTLA	O18745 antechinomys
13	45	50.6	61	1 HSP1_ANTSW	P42130 antechinus
14	45	50.6	61	1 HSP1_PARBI	O18768 parantechin
15	45	50.6	61	1 HSP1_SARHA	P42151 sarcophilus
16	45	50.6	62	1 HSP1_DASRO	P42134 dasykaluta
17	45	50.6	62	1 HSP1_MURLO	P42140 murexia lon
18	45	50.6	63	1 HSP1_ANTST	P42129 antechinus
19	44	49.4	140	1 VNB_PSV	P27336 lily symphy
20	43	48.3	63	1 HSP1_DROAU	P42132 dromiclops
21	43	48.3	135	1 LEG1_CONMY	P26788 conger myri
22	43	48.3	149	1 PGI_PIG	P49933 sus scrofa
23	43	48.3	413	1 FL1_TOBAC	Q40504 nicotiana t
24	42	47.2	191	1 YP6_AGRTU	P04030 agrobacteri
25	41	46.1	219	1 SFA_ARCFU	O28756 archaeoglob
26	40	44.9	416	1 FL2_TOBAC	Q40505 nicotiana t
27	40	44.9	794	1 YAFH_ECOLI	P47146 escherichia
28	40	44.9	1070	1 PVDF_PLAKN	Q50494 plasmodium
29	40	44.9	1147	1 DP3A_BORBU	O51526 borellia bu
30	40	44.9	3866	1 HRX_MOUSE	P59200 mus musculu
31	39	43.8	51	1 RL34_BORBU	P29220 borellia bu
32	39	43.8	3969	1 HRX_HUMAN	Q03164 homo sapien
33	38.5	43.3	464	1 IFE_BRALA	Q04948 branchiosto

RESULT 1

ID	PGI_PIG	STANDARD;	PRT;	149 AA.
AC	P32194;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	PROTEGRIN 1 PRECURSOR (PG-1) (NEUTROPHIL PEPTIDE 1).			
GN	NPGL			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow;			
RX	MEDLINE=94283613; PubMed=8013647;			
RA	Zhao C., Liu L., Lehrer R.I.;			
RT	"Identification of a new member of the protegrin family by cDNA cloning.";			
RL	FEBS Lett. 346:285-288(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=RED DUROC;			
RX	MEDLINE=95354835; PubMed=7628604;			
RA	Zhao C., Ganz T., Lehrer R.I.;			
RT	"The structure of porcine protegrin genes.";			
RL	FEBS Lett. 368:197-202(1995).			
RN	[3]			
RP	SEQUENCE OF 131-148.			
RC	TISSUE=Leukocyte;			
RX	MEDLINE=93327946; PubMed=8335113;			
RA	Kokryakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,			
RA	Aleshina G.M., Shamova O.V., Korneva H.A., Lehrer R.I.;			
RT	"Protegrins: leukocyte antimicrobial peptides that combine features of corticostatic defensins and tachyplesins.";			
RL	FEBS Lett. 327:231-236(1993).			
RN	[4]			
RP	SEQUENCE OF 131-148.			
RC	TISSUE=Neutrophils;			
RX	MEDLINE=93387466; PubMed=8375505;			
RA	Mirgorodskaya O.A., Shevchenko A.A., Abdalla K.O.M.A.,			
RA	Chernushevich I.V., Egorov T.A., Musoliamov A.X., Kokryakov V.N.,			
RT	"Primary structure of three cationic peptides from porcine neutrophils. Sequence determination by the combined usage of electrospray ionization mass spectrometry and Edman degradation.";			
RL	FEBS Lett. 330:339-342(1993).			
RN	[5]			
RP	STRUCTURE BY NMR OF PROTEGRIN 1.			
RX	MEDLINE=96235220; PubMed=8647100;			
RA	Aumelase A., Mantoni M., Roumestand C., Chiche L., Despaux E.,			
RA	Grassy G., Calas B., Chavanleu A.;			
RT	"Synthesis and solution structure of the antimicrobial peptide protegrin-1.";			
RL	Eur. J. Biochem. 237:575-583(1996).			

P16654 potato viru  
Q45486 staphylococ  
O95718 homo sapien  
O50715 mycobacteri  
P43675 e bifunctio  
Q98698 bacillus ha  
P36420 lactobacilli  
Q98705 xenopus lae  
P21238 arabidopsis  
P42145 pseudochiro  
O94142 conger myri  
P45952 mus musculu

[6]  
 RN STRUCTURE BY NMR OF PROTEGRIN 1.  
 RP MEDLINE=97113279; PubMed=8807886;  
 RA Fahrner R.L., Dieckmann T., Harwig S.S., Lehrer R.I., Eisenberg D.,  
 RA Feigon J.;  
 RT "Solution structure of protegrin-1, a broad-spectrum antimicrobial  
 peptide from porcine leukocytes.";  
 RL Chem. Biol. 3:543-550(1996).  
 CC -1- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA  
 CC MONOCYTOGENES AND C.ALBICANS, IN VITRO.  
 CC -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X79868; CAA56251.1; -;  
 DR EMBL; X84094; CAA58890.1; -;  
 DR PIR; S34585; S34585.  
 DR PIR; S36820; S36820.  
 DR PDB; 1PGI; 27-MAY-98.  
 DR InterPro; IPR001894; Cathelicidin.  
 DR Pfam; PF00666; Cathelicidins; 1.  
 DR ProDom; PD001838; Cathelicidins; 1.  
 DR PROSITE; PS00946; CATHELICIDINS\_1; 1.  
 DR PROSITE; PS00947; CATHELICIDINS\_2; 1.  
 DR Antibiotic; Signal; Amidation; Multigene family; 3D-structure.  
 KW SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 130  
 FT CHAIN 131 148  
 FT MOD\_RES 30 30  
 FT DISULFID 85 96  
 FT DISULFID 107 124  
 FT DISULFID 136 145  
 FT DISULFID 138 143  
 FT MOD\_RES 148 148  
 FT SEQUENCE 149 AA; 16677 MW; 6EFBA98429CD6EC4 CRC64;  
 SQ  
 Query Match 77.5%; Score 69; DB 1; Length 149;  
 Best Local Similarity 77.8%; Pred. No. 3e-05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 RGGRLSYRRRFFSVSGR 18  
 Db 131 RGGRLCYRRRFFCVGVR 148  
 RESULT 2  
 PG2\_PIG  
 ID PG2\_PIG STANDARD; PRT; 147 AA.  
 AC P32195;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROTEGRIN 2 PRECURSOR (PG-2).  
 GN NPG2.  
 OS Sus scrofa (pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=94071898; PubMed=8250892;  
 RA Storici P., Zanetti M.;  
 RT "A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide  
 with a cathelin-like pro-sequence.";

Biochem. Biophys. Res. Commun. 196:1363-1368(1993).  
 [2]  
 RN SEQUENCE OF 131-146.  
 RX TISSUE=Leukocyte;  
 CC MEDLINE=9337946; PubMed=8335113;  
 RA Kokryakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,  
 RA Aleshina G.M., Shamova O.V., Korneva H.A., Lehrer R.I.;  
 RT "Protegrins: leukocyte antimicrobial peptides that combine features  
 of corticostatic defensins and tachyplesins.";  
 RL FEBS Lett. 327:231-236(1993).  
 CC -1- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA  
 CC MONOCYTOGENES AND C.ALBICANS, IN VITRO.  
 CC -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L24745; AAA31061.1; -;  
 DR HSP; P32194; 1PGI.  
 DR InterPro; IPR001894; Cathelicidin.  
 DR Pfam; PF00666; Cathelicidins; 1.  
 DR ProDom; PD001838; Cathelicidins; 1.  
 DR PROSITE; PS00946; CATHELICIDINS\_1; 1.  
 DR PROSITE; PS00947; CATHELICIDINS\_2; 1.  
 DR Antibiotic; Signal; Amidation; Multigene family.  
 KW SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 130  
 FT CHAIN 131 146  
 FT MOD\_RES 30 30  
 FT DISULFID 85 96  
 FT DISULFID 107 124  
 FT DISULFID 136 145  
 FT DISULFID 138 143  
 FT MOD\_RES 146 146  
 FT SEQUENCE 147 AA; 16478 MW; 698429DFEC40466 CRC64;  
 SQ  
 Query Match 70.8%; Score 63; DB 1; Length 147;  
 Best Local Similarity 70.6%; Pred. No. 0.00035;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 RGGRLSYRRRFFSVSG 17  
 Db 131 RGGRLCYRRRFFCICVG 147  
 RESULT 3  
 PG3\_PIG  
 ID PG3\_PIG STANDARD; PRT; 149 AA.  
 AC P32196;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROTEGRIN 3 PRECURSOR (PG-3).  
 GN NPG3.  
 OS Sus scrofa (pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=94283613; PubMed=8013647;  
 RA Zhao C., Liu L., Lehrer R.I.;  
 RT "Identification of a new member of the protegrin family by cDNA  
 cloning.";  
 RL FEBS Lett. 346:285-288(1994).

```

[2]
RN  SEQUENCE FROM N.A.
RP  STRAIN-RED DUROC;
RX  MEDLINE=95354835; PubMed=7628604;
RA  Zhao C., Ganz T., Lehrer R.I.;
RT  "The structure of porcine proteogrin genes.";
RL  FEBS Lett. 368:197-202(1995).
[3]
RN  SEQUENCE OF 131-148.
RP  TISSUE=Leukocyte;
RX  MEDLINE=93327946; PubMed=8335113;
RA  Kokiyakov V.N., Harwig S.S.L., Panyutich E.A., Shevchenko A.A.,
RA  Aleshina G.M., Shamova O.V., Korneva H.A., Lehrer R.I.;
RT  "Protegrins: leukocyte antimicrobial peptides that combine features
RT  of corticostatic defensins and tachyplesins.";
RL  FEBS Lett. 327:231-236(1993).
CC  -1- FUNCTION: MICROBICIDAL ACTIVITY. ACTIVE AGAINST E.COLI, LISTERIA
CC  MONOCYTOGENES AND C.ALBCANS, IN VITRO.
CC  -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; X83267; CAA58240.1; -
DR  EMBL; X84095; CAA58891.1; -
DR  PIR; S34587; S34587.
DR  HSP; P32194; LPGL.
DR  InterPro; IPR001894; Cathelicidin.
DR  Pfam; PF00666; Cathelicidins; 1.
DR  ProDom; PD001838; Cathelicidin; 1.
DR  PROSITE; PS00946; CATHELICIDINS_1; 1.
DR  PROSITE; PS00947; CATHELICIDINS_2; 1.
DR  Antibiotic; Signal; Amidation; Multigene family.
KW  SIGNAL 1 29 POTENTIAL.
FT  PROPEP 30 130
FT  CHAIN 131 148
FT  MOD_RES 30 30
FT  MOD_RES 30 30
FT  PYRROLIDONE CARBOXYLIC ACID (BY
FT  SIMILARITY).
FT  DISULFID 85 96
FT  DISULFID 107 124
FT  DISULFID 136 145
FT  DISULFID 138 143
FT  DISULFID 138 143
FT  MOD_RES 148 148
FT  AMIDATION (G-149 PROVIDE AMIDE GROUP).
FT  SEQUENCE 149 AA; 16578 MW; 6F4BA98429CD6ED4 CRC64;
-----
Query Match 69.7%; Score 62; DB 1; Length 149;
Best Local Similarity 72.2%; Pred. NO. 0.00053;
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 RGGRLSYRRRRFSVSVGR 18
Db 131 RGGGLCYCRRRCVCVGR 148
-----
RESULT 4
VNB_PVMG 4
ID PG5_PIG STANDARD; PRT; 149 AA.
AC P49934;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEGRIN 5 PRECURSOR (PG-5).
GN NPG5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Sulina; Suidae; Sus.
OX NCBI_TaxID=9823;
-----

```

```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN-RED DUROC;
RX  MEDLINE=95354835; PubMed=7628604;
RA  Zhao C., Ganz T., Lehrer R.I.;
RT  "The structure of porcine proteogrin genes.";
RL  FEBS Lett. 368:197-202(1995).
CC  -1- FUNCTION: MICROBICIDAL ACTIVITY (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; X84096; CAA58892.1; -
DR  HSP; P32194; LPGL.
DR  InterPro; IPR001894; Cathelicidin.
DR  Pfam; PF00666; Cathelicidins; 1.
DR  ProDom; PD001838; Cathelicidin; 1.
DR  PROSITE; PS00946; CATHELICIDINS_1; 1.
DR  PROSITE; PS00947; CATHELICIDINS_2; 1.
DR  Antibiotic; Amidation; Multigene family; Signal.
KW  SIGNAL 1 29 POTENTIAL.
FT  PROPEP 30 130
FT  CHAIN 131 148
FT  MOD_RES 30 30
FT  MOD_RES 30 30
FT  PYRROLIDONE CARBOXYLIC ACID (BY
FT  SIMILARITY).
FT  DISULFID 85 96
FT  DISULFID 107 124
FT  DISULFID 136 145
FT  DISULFID 138 143
FT  DISULFID 138 143
FT  MOD_RES 148 148
FT  AMIDATION (G-149 PROVIDE AMIDE GROUP)
FT  (BY SIMILARITY).
FT  SEQUENCE 149 AA; 16604 MW; 6CC7262429CD6B64 CRC64;
-----
Query Match 69.7%; Score 62; DB 1; Length 149;
Best Local Similarity 72.2%; Pred. NO. 0.00053;
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 RGGRLSYRRRRFSVSVGR 18
Db 131 RGGRLCYCRRRCVCVGR 148
-----
RESULT 5
VNB_PVMG 5
ID VNB_PVMG STANDARD; PRT; 108 AA.
AC Q01687;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE 12 KDA PROTEIN (PR12) (NUCLEIC ACID-BINDING REGULATORY PROTEIN).
OS Potato virus M (German isolate) (PVM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=31710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91092429; PubMed=2265707;
RA Gramstat A., Courtzopoulos A., Rohde W.;
RT "The 12 kDa protein of potato virus M displays properties of a
RT nucleic acid-binding regulatory protein.";
RL FEBS Lett. 276:34-38(1990).
CC -1- FUNCTION: MAY ACT AS A REGULATORY FACTOR DURING VIRAL
CC TRANSCRIPTION.
CC -1- SIMILARITY: TO NUCLEIC ACID-BINDING PROTEINS FROM OTHER
CC CARLAVIRUSES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X57440; CAA040689.1; -  
DR PIR; S12976; S12976.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
KW Zinc-finger; DNA-binding.  
FT ZN\_FING 57 78 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 108 AA; 12119 MW; 86E80F8DC0B376E8 CRC64;

Query Match 57.3%; Score 51; DB 1; Length 108;  
Best Local Similarity 52.9%; Pred. No. 0.033;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRFSVSVGR 18  
||| |::| |::| |  
Db 40 GGRSKYARRRAISIAIR 56

RESULT 6  
VNBP\_PVNR STANDARD; PRT; 108 AA.  
AC P17530;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE 10.7 KDA PROTEIN (PUTATIVE NUCLEIC ACID-BINDING PROTEIN).  
OS Potato virus M (strain Russian) (PVM).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=12168;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89293091; PubMed=2738581;  
RA Rupakov V.V., Morozov S.Y., Kanyuka K.V., Zavriev S.K.;  
RT "Partial nucleotide sequence of potato virus M RNA shows similarities  
RT to proteoviruses in gene arrangement and the encoded amino acid  
RT sequences".  
RL J. Gen. Virol. 70:1861-1869(1989).  
RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=91116326; PubMed=1990070;  
RA Zavriev S.K., Kanyuka K.V., Levay K.E.;  
RT "The genome organization of potato virus M RNA.";  
RL J. Gen. Virol. 72:9-14(1991).  
FT ZN\_FING 57 78 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 107 AA; 12576 MW; 674D16319920ED2F CRC64;

Query Match 52.8%; Score 47; DB 1; Length 107;  
Best Local Similarity 56.2%; Pred. No. 0.17;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GRLSYRRRFSVSVGR 18  
||| |::| |::| |  
Db 40 GRSYARRRALELGR 55

RESULT 8  
HSP1\_MACRU STANDARD; PRT; 59 AA.  
AC P42142;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Macropus rufus (Red kangaroo) (Megalania rufa).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D14449; BAA03344.1; -  
DR EMBL; X53062; CAA37237.1; -  
DR PIR; PNO006; WMVP5.  
DR PIR; S21606; S21606.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
KW Zinc-finger; DNA-binding.  
FT ZN\_FING 57 78  
FT CONFLICT 80 108 C4-TYPE (POTENTIAL).  
PGISYNNRVAQFIDEGVTEVPVINKRE ->  
LVSLTMCARRLLMKE (IN REF. 1).

SQ SEQUENCE 108 AA; 12183 MW; 00886E246A553B53 CRC64;

Query Match 57.3%; Score 51; DB 1; Length 108;  
Best Local Similarity 52.9%; Pred. No. 0.033;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRFSVSVGR 18  
||| |::| |::| |  
Db 40 GGRSKYARRRAISIAIR 56

RESULT 7  
VNBP\_CVB STANDARD; PRT; 107 AA.  
ID VNBP\_CVB  
AC P37992;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE 12.6 KDA PROTEIN (PUTATIVE NUCLEIC ACID-BINDING PROTEIN).  
OS Chrysanthemum virus B (CVB).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
OX NCBI\_TaxID=12165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92013948; PubMed=1919520;  
RA Levay K., Zavriev S.;  
RT "Nucleotide sequence and gene organization of the 3'-terminal region  
RT of chrysanthemum virus B genomic RNA".  
RL J. Gen. Virol. 72:2333-2337(1991).  
CC -1- FUNCTION: MAY ACT AS A REGULATORY FACTOR DURING VIRAL  
CC TRANSCRIPTION.  
CC -1- SIMILARITY: TO NUCLEIC ACID-BINDING PROTEINS FROM OTHER  
CC CARLAVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; S60150; AAB20081.1; -  
DR PIR; JQ1251; JQ1251.  
DR InterPro; IPR002568; Carla\_C4.  
DR Pfam; PF01623; Carla\_C4; 1.  
KW Zinc-finger; DNA-binding.  
FT ZN\_FING 56 78 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 107 AA; 12576 MW; 674D16319920ED2F CRC64;

Query Match 52.8%; Score 47; DB 1; Length 107;  
Best Local Similarity 56.2%; Pred. No. 0.17;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GRLSYRRRFSVSVGR 18  
||| |::| |::| |  
Db 40 GRSYARRRALELGR 55

RESULT 8  
HSP1\_MACRU STANDARD; PRT; 59 AA.  
AC P42142;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Macropus rufus (Red kangaroo) (Megalania rufa).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.



```

Db 42 RRRRGYSRRYS 54

RESULT 11
HSP1_MACAG
ID HSP1_MACAG STANDARD; PRT; 60 AA.
AC P42137;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Macropus agilis (Agile wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9313;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SpERM;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L35451; AAA74615.1; -
CC InterPro; IPR000221; Protamine_P1.
CC Pfam; PF00260; protamine_P1; 1.
CC PROSITE; PS00048; PROTAMINE_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
CC Testis; DNA condensation; Nuclear protein.
CC INIT_MET 0 BY SIMILARITY.
CC SEQUENCE 60 AA; 8338 MW; 96255C818921EB85 CRC64;
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L35451; AAA74615.1; -
CC InterPro; IPR000221; Protamine_P1.
CC Pfam; PF00260; protamine_P1; 1.
CC PROSITE; PS00048; PROTAMINE_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
CC Testis; DNA condensation; Nuclear protein.
CC INIT_MET 0 BY SIMILARITY.
CC SEQUENCE 60 AA; 8338 MW; 96255C818921EB85 CRC64;

Query Match 50.6%; Score 45; DB 1; Length 60;
Best Local Similarity 69.2%; Pred. No. 0.2;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RGRRLYSRRRFS 13
||| |||||:|
Db 41 RRRRGYSRRYS 53

RESULT 12
HSP1_ANTLA
ID HSP1_ANTLA STANDARD; PRT; 61 AA.
AC O18745;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Antechinus laniger.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinomys.
OX NCBI_TaxID=60701;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97446280; PubMed=9299728;
RA Krajewski C., Blacket M., Buckley L., Westernman M.;

Query Match 50.6%; Score 45; DB 1; Length 60;
Best Local Similarity 69.2%; Pred. No. 0.2;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RGRRLYSRRRFS 13
||| |||||:|
Db 41 RRRRGYSRRYS 53

RESULT 13
HSP1_ANTSW
ID HSP1_ANTSW STANDARD; PRT; 61 AA.
AC P42130; P42146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Antechinus swainsonii, Phascosorex dorsalis,
OS Neophascogale lorentzii (Long-clawed marsupial mouse),
OS Dasyurus albopunctatus (Native cat),
OS Dasyurus geoffroyi (Chuditch/western quoll), and
OS Dasyurus spartacus (Native cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.
OX NCBI_TaxID=9284, 9295, 32551, 32545, 63143, 32546;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=A.swainsonii, and P.dorsalis; TISSUE=SpERM;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=N.lorentzii, D.albopunctatus, D.geoffroyi, and D.spartacus;
RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
RT cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----

```

```

RT "A multigene assessment of phylogenetic relationships within the
RT dasyurid marsupial subfamily Sminthopsinae.";
RL Mol. Phylogenet. Evol. 8:236-248(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001587; AAB91377.1; -
CC InterPro; IPR000221; Protamine_P1.
CC Pfam; PF00260; protamine_P1; 1.
CC PROSITE; PS00048; PROTAMINE_P1; 1.
CC Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
CC Testis; DNA condensation; Nuclear protein.
CC INIT_MET 0 BY SIMILARITY.
CC SEQUENCE 61 AA; 8409 MW; E021567627E562B1 CRC64;
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC Query Match 50.6%; Score 45; DB 1; Length 61;
CC Best Local Similarity 69.2%; Pred. No. 0.21;
CC Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RGRRLYSRRRFS 13
||| |||||:|
Db 42 RRRRGYSRRYS 54

RESULT 13
HSP1_ANTSW
ID HSP1_ANTSW STANDARD; PRT; 61 AA.
AC P42130; P42146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Antechinus swainsonii, Phascosorex dorsalis,
OS Neophascogale lorentzii (Long-clawed marsupial mouse),
OS Dasyurus albopunctatus (Native cat),
OS Dasyurus geoffroyi (Chuditch/western quoll), and
OS Dasyurus spartacus (Native cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.
OX NCBI_TaxID=9284, 9295, 32551, 32545, 63143, 32546;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=A.swainsonii, and P.dorsalis; TISSUE=SpERM;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=N.lorentzii, D.albopunctatus, D.geoffroyi, and D.spartacus;
RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
RT cytochrome b, 12S rRNA, and protamine P1 gene trees.";
RL J. Mammal. Evol. 4:217-236(1997).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----

```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; L35338; AAB95429.1; -;  
 DR EMBL; L35339; AAA74601.1; -;  
 DR EMBL; L35339; AAA74601.1; -;  
 DR EMBL; AF010267; AAB69297.1; -;  
 DR EMBL; AF010272; AAB89302.1; -;  
 DR EMBL; AF010274; AAB69304.1; -;  
 DR EMBL; AF010275; AAB69305.1; -;  
 DR InterPro; IPR000221; Protamine\_P1.  
 DR Pfam; PF00260; Protamine\_P1; 1.  
 DR PROSITE; PS00048; Protamine\_P1; 1.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 FT Testis; DNA condensation; Nuclear protein.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 SQ SEQUENCE 61 AA; 8390 MW; E021472785E71221 CRC64;

Query Match 50.6%; Score 45; DB 1; Length 61;  
 Best Local Similarity 69.2%; Pred. No. 0.21;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGRRLSYRRRFS 13  
 II | | | | | | |  
 DB 42 RGRRCYSRRYS 54

RESULT 14  
 HSPL\_PARB1 STANDARD; PRT; 61 AA.  
 AC O18768;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS Parantechnus bilarni (Broad-footed marsupial mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Parantechninus.  
 OX NCBI\_TaxID=32555;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Krajewski C., Young J., Buckley L., Woolley P.A., Westerman M.;  
 RT "Reconstructing the taxonomic radiation of dasyurine marsupials with  
 RT cytochrome b, 12S rRNA, and protamine P1 gene trees.";  
 RL J. Mammal. Evol. 4:217-236(1997).  
 CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- TISSUE SPECIFICITY: TESTIS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AF010277; AAB69307.1; -;  
 DR InterPro; IPR000221; Protamine\_P1.  
 DR Pfam; PF00260; Protamine\_P1; 1.  
 DR PROSITE; PS00048; Protamine\_P1; 1.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 FT Testis; DNA condensation; Nuclear protein.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 SQ SEQUENCE 61 AA; 8421 MW; C02857DF087FC9A9 CRC64;

Query Match 50.6%; Score 45; DB 1; Length 61;  
 Best Local Similarity 69.2%; Pred. No. 0.21;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGRRLSYRRRFS 13  
 II | | | | | | |  
 DB 43 RGRRCYSRRYS 55

RESULT 15  
 HSPL\_SARHA STANDARD; PRT; 61 AA.  
 AC P42151;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SPERM PROTAMINE P1.  
 GN PRM1.  
 OS Sarcophilus harrisii (Tasmanian devil), and  
 OS Dasyurus maculatus (Tiger quoll).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sarcophilus.  
 OX NCBI\_TaxID=9305, 9281;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.harrisii; TISSUE=Sperm;  
 RX MEDLINE=95215351; PubMed=7700877;  
 RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
 RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL Proc. F. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=D.maculatus;  
 RA Krajewski C., Young J., Buckley L., Woolley P.A., Westerman M.;  
 RT "Reconstructing the taxonomic radiation of dasyurine marsupials with  
 RT cytochrome b, 12S rRNA, and protamine P1 gene trees.";  
 RL J. Mammal. Evol. 4:217-236(1997).  
 CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- TISSUE SPECIFICITY: TESTIS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; L35324; AAA74608.1; -;  
 DR EMBL; AF010276; AAB69306.1; -;  
 DR InterPro; IPR000221; Protamine\_P1.  
 DR Pfam; PF00260; Protamine\_P1; 1.  
 DR PROSITE; PS00048; Protamine\_P1; 1.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 FT Testis; DNA condensation; Nuclear protein.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 SQ SEQUENCE 61 AA; 8410 MW; 4A215D3D85E71230 CRC64;

Query Match 50.6%; Score 45; DB 1; Length 61;  
 Best Local Similarity 69.2%; Pred. No. 0.21;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RGRRLSYRRRFS 13  
 II | | | | | | |  
 DB 42 RGRRCYSRRYS 54

Search completed: February 12, 2002, 12:39:48  
Job time: 801 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:25:22 ; Search time 126.85 seconds  
(without alignments)  
10.809 Million cell updates/sec

Title: US-09-485-571-15

Perfect score: 89

Sequence: 1 RGRSLYSRRRFSVSGR 18:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	77.5	149	2 S57607	protegrin 1 precu
2	63	70.8	147	2 JN0900	protegrin 2 precu
3	62	69.7	149	2 S57609	protegrin 5 precu
4	62	69.7	149	2 A53895	protegrin 3 precu
5	51	57.3	108	1 WMVVP5	nucleic acid-bind
6	51	57.3	108	2 S12976	12K protein - pota
7	49	55.1	94	1 B48549	nucleic acid-bind
8	47	52.8	107	2 JQ1251	hypothetical 12.6K
9	44	49.4	627	2 G96537	hypothetical prote
10	43	48.3	135	2 S21102	lectin - eel (Cong
11	43	48.3	149	2 B53895	protegrin 4 precu
12	43	48.3	413	2 T03240	FLO/LFY protein ho
13	43	48.3	1016	2 T30942	aminopeptidase (EC
14	43	48.3	1016	2 T30943	aminopeptidase (EC
15	42	47.2	118	2 C72642	hypothetical prote
16	42	47.2	191	1 QQA66T	hypothetical prote
17	41	46.1	186	2 A30832	hypothetical prote
18	41	46.1	219	2 C69439	sugar fermentation
19	41	46.1	447	2 S53982	hypothetical prote
20	41	46.1	1696	2 T00057	hypothetical prote
21	40	44.9	134	2 D81096	hypothetical prote
22	40	44.9	349	2 T06680	hypothetical prote
23	40	44.9	416	2 T03243	hypothetical prote
24	40	44.9	416	2 D71936	FLO/LFY protein ho
25	40	44.9	747	2 T42599	hypothetical prote
26	40	44.9	826	2 F85510	minor capsid prote
27	40	44.9	826	2 F64746	probable acyl-CoA
28	40	44.9	1030	2 T16114	probable membrane
29	40	44.9	1161	2 B70172	DNA polymerase III

30	40	44.9	3869	2 A48205	All-1 protein +GTE
31	39.5	44.4	187	2 B72545	probable tRNA intr
32	39	43.8	51	2 G70154	ribosomal protein
33	39	43.8	167	2 E69554	conserved hypothet
34	39	43.8	201	2 B72739	hypothetical prote
35	39	43.8	217	2 T33652	hypothetical prote
36	39	43.8	310	2 T43147	hypothetical prote
37	39	43.8	325	2 T38308	hypothetical prote
38	39	43.8	349	2 S55626	hypothetical prote
39	39	43.8	532	2 H72730	probable acyl-CoA
40	39	43.8	647	2 A84265	hypothetical prote
41	39	43.8	760	2 D82164	oxidoreductase, ac
42	39	43.8	1028	2 G64595	acrilflavine resist
43	39	43.8	1224	2 T26377	hypothetical prote
44	39	43.8	2114	2 E96505	hypothetical prote
45	39	43.8	3968	2 A44265	trithorax homolog

ALIGNMENTS

RESULT 1  
S57607  
protegrin 1 precursor - pig  
N:Alternate names: neutrophil peptide 1  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S66284; S45712; S36820; S34585; S57607  
R:Zhao, C.; Ganz, T.; Lehrer, R.I.  
FEBS Lett. 368, 197-202, 1995  
A:Title: The structure of porcine protegrin genes.  
A:Reference number: S66283; MUID:95354835  
A:Accession: S66284  
A:Molecule type: DNA  
A:Residues: 1-149 <ZHA>  
A:Cross-references: EMBL:X84094; NID:9887642; PIDN:CAA58890.1; PID:9887643  
R:Zhao, C.; Liu, L.; Lehrer, R.I.  
FEBS Lett. 346, 285-288, 1994  
A:Title: Identification of a new member of the protegrin family by cDNA cloning.  
A:Reference number: S45712; MUID:94283613  
A:Accession: S45712  
A:Molecule type: mRNA  
A:Residues: 1-149 <ZH2>  
A:Cross-references: GB:X79868; NID:9603035; PIDN:CAA56251.1; PID:9603036  
R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Egr  
FEBS Lett. 330, 339-342, 1993  
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Seq  
A:Reference number: S36820; MUID:93387466  
A:Accession: S36820  
A:Molecule type: protein  
A:Residues: 131-148 <MIR>  
R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M  
FEBS Lett. 327, 231-236, 1993  
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of cort  
A:Reference number: S34585; MUID:93327946  
A:Accession: S34585  
A:Molecule type: protein  
A:Residues: 131-148 <KOK>  
C:Genetics:  
A:Gene: NPG1  
A:Introns: 66/3; 102/3; 126/3  
C:Superfamily: cathelin; cystatin homology  
C:Keywords: amidated carboxyl end; antibacterial; neutrophil  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:22-129/Domain: cystatin homology <CYS>  
F:30-130/Domain: propeptide #status predicted <PRO>  
F:131-148/Product: protegrin 1 #status experimental <MAT>  
F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followi

Query Match 77.5%; Score 69; DB 2; Length 149;  
Best Local Similarity 77.8%; Pred. No. 0.00022;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

Qy 1 RGRSLYSRRRFSVSVGR 18
      ||||| | |||| | |||
Db 131 RGRRLCYCRRCFCVGR 148

RESULT 2
JN0900
protegrin 2 precursor - pig
N:Alternate names: cathelin-like protein precursor; neutrophil peptide 3
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C:Accession: JN0900; S36822; S34586
R:Storici, P.; Zanetti, M.
Biochem. Biophys. Res. Commun. 196, 1363-1368, 1993
A:Title: A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide with a cat
A:Reference number: JN0900; MUID:94071898
A:Accession: JN0900
A:Molecule type: mRNA
A:Residues: 1-147 <SIG>
R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Egorov,
FEBS Lett. 330, 339-342, 1993
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Sequence
A:Reference number: S36820; MUID:93387466
A:Accession: S36822
A:Molecule type: protein
A:Residues: 131-146 <MIR>
R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.; Sh
FEBS Lett. 327, 231-236, 1993
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of corticost
A:Reference number: S34585; MUID:93327946
A:Accession: S34586
A:Molecule type: protein
A:Residues: 131-146 <KOK>
C:Comment: This protein plays a role in the nonoxidative antimicrobial defense mechanism
C:Superfamily: cathelin; cystatin homology
C:Keywords: amidated carboxyl end; antibacterial; neutrophil
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:131-146/Product: protegrin 2 #status experimental <MAT>
F:146/Modified site: amidated carboxyl end (Val) (amide in mature form from following gl

Query Match 70.8%; Score 63; DB 2; Length 147;
Best Local Similarity 70.6%; Pred. No. 0.0022;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGRSLYSRRRFSVSVG 17
      ||||| | |||| | ||
Db 131 RGRRLCYCRRCFCVGR 147

RESULT 3
S57609
protegrin 5 precursor - pig
N:Alternate names: cathelin-associated antimicrobial peptide
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 16-Jul-1999
C:Accession: S56283; S57609
R:Zhao, C.; Ganz, T.; Lehrer, R.I.
FEBS Lett. 368, 197-202, 1995
A:Title: The structure of porcine protegrin genes.
A:Reference number: S56283; MUID:95354835
A:Accession: S56283
A:Molecule type: DNA
A:Residues: 1-149 <ZHA>
A:Cross-references: EMBL:X84096; NID:9887646; PIDN:CAA58892.1; PID:9887647
A:Experimental source: leukocytes
C:Genetics:
A:Gene: NPG5
A:Introns: 66/3; 102/3; 126/3
C:Superfamily: cathelin; cystatin homology

Qy 1 RGRSLYSRRRFSVSVGR 18
      ||||| | |||| | |||

```

```

C:Keywords: amidated carboxyl end; antibacterial
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:30-130/Domain: propeptide #status predicted <PRO>
F:131-148/Product: protegrin 5 #status predicted <MAT>
F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followin

Query Match 69.7%; Score 62; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 0.0033;
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RGRSLYSRRRFSVSVGR 18
      ||||| | |||| | |||
Db 131 RGRRLCYCRRCFCVGR 148

RESULT 4
A53895
protegrin 3 precursor - pig
N:Alternate names: neutrophil peptide 2
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C:Accession: S56285; A53895; S34587; S36821; S57608
R:Zhao, C.; Ganz, T.; Lehrer, R.I.
FEBS Lett. 368, 197-202, 1995
A:Title: The structure of porcine protegrin genes.
A:Reference number: S56283; MUID:95354835
A:Accession: S56285
A:Molecule type: DNA
A:Residues: 1-149 <ZH3>
A:Cross-references: EMBL:X84095; NID:9887644; PIDN:CAA58891.1; PID:9887645
R:Zhao, C.; Liu, L.; Lehrer, R.I.
FEBS Lett. 346, 285-288, 1994
A:Title: Identification of a new member of the protegrin family by cDNA cloning.
A:Reference number: S45712; MUID:94283613
A:Accession: A53895
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-149 <ZHA>
A:Cross-references: GB:X83267; NID:g603037; PIDN:CAA58240.1; PID:g603038
R:Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.
FEBS Lett. 327, 231-236, 1993
A:Title: Protegrins: leukocyte antimicrobial peptides that combine features of corti
A:Reference number: S34585; MUID:93327946
A:Accession: S34587
A:Molecule type: protein
A:Residues: 131-148 <KOK>
R:Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Ego
FEBS Lett. 330, 339-342, 1993
A:Title: Primary structure of three cationic peptides from porcine neutrophils. Sequ
A:Reference number: S36820; MUID:93387466
A:Accession: S36821
A:Molecule type: protein
A:Residues: 131-148 <MIR>
C:Genetics:
A:Gene: NPG3
A:Introns: 66/3; 102/3; 126/3
C:Superfamily: cathelin; cystatin homology
C:Keywords: amidated carboxyl end; antibacterial; neutrophil
F:1-29/Domain: signal sequence #status predicted <SIG>
F:22-129/Domain: cystatin homology <CYS>
F:30-130/Domain: propeptide #status predicted <PRO>
F:131-148/Product: protegrin 3 #status experimental <MAT>
F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followin

Query Match 69.7%; Score 62; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 0.0033;
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RGRSLYSRRRFSVSVGR 18
      ||||| | |||| | |||

```

Db 131 RGGGLCYCRRRCVCVGR 148

RESULT 5  
MWVYP5  
nucleic acid-binding protein - potato virus M (strain Russian)  
C;Species: potato virus M  
A;Note: host Lycopersicon esculentum (tomato)  
C;Date: 31-Mar-1990 #sequence\_revision 23-Mar-1995 #text\_change 29-Oct-1999  
C;Accession: F54333; PNO006; PNO095; S21606  
R;Zavriev, S.K.; Kanyuka, K.V.; Levay, K.E.  
J. Gen. Virol. 72, 9-14, 1991  
A;Title: The genome organization of potato virus M RNA.  
A;Reference number: A54333; MUID:91116326  
A;Accession: F54333  
A;Molecule type: genomic RNA  
A;Residues: 1-108 <ZAV>  
A;Cross-references: EMBL:X53062; NID:g61291; PIDN:CAA37237.1; PID:g61297  
R;Rupakov, V.V.; Morozov, S.Y.; Kanyuka, K.V.; Levay, S.K.  
J. Gen. Virol. 70, 1861-1869, 1989  
A;Title: Partial nucleotide sequence of potato virus M RNA shows similarities to potexv  
A;Reference number: A92800; MUID:89293091  
A;Accession: PNO006  
A;Molecule type: mRNA  
A;Residues: 1-79, 'LVSLTMCARNLMLKE' <RUP>  
A;Note: this sequence has been corrected  
R;Zavriev, S.K.; Kanyuka, K.V.; Levay, K.E.  
Mol. Biol. (Mosk.) 25, 761-769, 1991  
A;Title: The complete nucleotide sequence of potato virus M genomic RNA.  
A;Reference number: PNO093; MUID:92049299  
A;Accession: PNO095  
A;Molecule type: genomic RNA  
A;Residues: 1-108 <ZA2>  
A;Cross-references: GB:X53062; NID:g61291; PIDN:CAA37237.1; PID:g61297  
A;Note: this is a revision to the sequence from reference A92800  
R;Zavriev, S.K.  
submitted to the EMBL Data Library, May 1990  
A;Reference number: S21601  
A;Accession: S21606  
A;Status: preliminary  
A;Molecule type: genomic RNA  
A;Residues: 1-79, 'LVSLTMCARNLMLKE' <ZA3>  
A;Cross-references: EMBL:X53062  
A;Experimental source: Russian wild type  
C;Superfamily: potato virus nucleic acid-binding protein  
C;Keywords: DNA binding; metal binding; nucleotide binding; zinc finger  
F;57-78/Region: zinc finger

Query Match 57.3%; Score 51; DB 1; Length 108;  
Best Local Similarity 52.9%; Pred. No. 0.18;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRFRFSVSVGR 18  
||| 1:|||| :||  
Db 40 GGRSKYARRRRRAISIAI 56

RESULT 6  
S12976  
12K protein - potato virus M  
C;Species: potato virus M  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Sep-1999  
C;Accession: S12976  
R;Gramstat, A.; Courtpozanis, A.; Rohde, W.  
FEBS Lett. 276, 34-38, 1990  
A;Title: The 12 kDa protein of potato virus M displays properties of a nucleic acid-binding  
A;Reference number: S12975; MUID:91092429  
A;Accession: S12976  
A;Status: preliminary  
A;Molecule type: genomic RNA  
A;Residues: 1-108 <GRA>  
A;Cross-references: GB:X57440; NID:g61400; PIDN:CAA40689.1; PID:g61402

C;Superfamily: potato virus nucleic acid-binding protein

Query Match 57.3%; Score 51; DB 2; Length 108;  
Best Local Similarity 52.9%; Pred. No. 0.18;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRFRFSVSVGR 18  
||| 1:|||| :||  
Db 40 GGRSKYARRRRRAISIAI 56

RESULT 7  
B48549  
nucleic acid-binding protein - potato virus S  
C;Species: potato virus S  
C;Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 23-Jul-1999  
C;Accession: B48549  
R;Foster, G.D.; Mills, P.R.  
Virus Genes 6, 213-220, 1992  
A;Title: The 3'-nucleotide sequence of an ordinary strain of potato virus S.  
A;Reference number: A48549; MUID:93033173  
A;Accession: B48549  
A;Molecule type: genomic RNA  
A;Residues: 1-94 <FOS>  
A;Cross-references: GB:S45593; NID:g256417; PIDN:AAB23462.1; PID:g256419  
A;Note: sequence extracted from NCBI backbone (NCBIN:114637, NCBIP:114639)  
C;Superfamily: potato virus nucleic acid-binding protein  
C;Keywords: DNA binding; zinc finger  
F;55-75/Region: zinc finger

Query Match 55.1%; Score 49; DB 1; Length 94;  
Best Local Similarity 52.9%; Pred. No. 0.34;  
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GGRLSYRRRFRFSVSVGR 18  
||| 1:|||| :||  
Db 38 GGRSTYARKRRARSIGR 54

RESULT 8  
JQ1251  
hypothetical 12.6K protein - chrysanthemum virus B  
C;Species: chrysanthemum virus B  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Sep-1999  
C;Accession: JQ1251  
R;Levay, K.; Zavriev, S.  
J. Gen. Virol. 72, 2333-2337, 1991  
A;Title: Nucleotide sequence and gene organization of the 3'-terminal region of chr

A;Reference number: JQ1246; MUID:92013948  
A;Accession: JQ1251  
A;Molecule type: genomic RNA  
A;Residues: 1-107 <LEV>

A;Cross-references: GB:S60150; NID:g237315; PIDN:AAB20081.1; PID:g237321  
C;Superfamily: potato virus nucleic acid-binding protein

Query Match 52.8%; Score 47; DB 2; Length 107;  
Best Local Similarity 56.2%; Pred. No. 0.83;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GRLSYRRRFRFSVSVGR 18  
||| 1:|||| :||  
Db 40 GRSSYARRRRRALELGR 55

RESULT 9  
G96537  
hypothetical protein F2J10.1 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: G96537

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: G96537  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-627 <STO>  
A:CROSS-references: GB:AE005173; NID:g8569089; PIDN:AAF76434.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F2J10.1  
A:Map position: 1

Query Match 49.4%; Score 44; DB 2; Length 627;  
Best Local Similarity 44.4%; Pred. No. 14;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 RGRLSYRRRFSVSVGR 18  
:||||| : : : :  
Db 279 KGRLSLPRESLEISTAR 296

RESULT 10  
S21102  
lectin - eel (Conger myriaster)  
C:Species: Conger myriaster  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 26-Aug-1999  
C:Accession: S21102  
R:Muramoto, K.; Kamiya, H.  
Biochim. Biophys. Acta 1116, 129-136, 1992  
A:Title: The amino-acid sequence of a lectin from conger eel, Conger myriaster, skin muc  
A:Reference number: S21102; MUID:92256465  
A:Accession: S21102  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-135 <MUR>  
C:Superfamily: beta-galactoside-binding lectin

Query Match 48.3%; Score 43; DB 2; Length 135;  
Best Local Similarity 56.2%; Pred. No. 4.9;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFSVSVG 17  
|| : : | : ||| : ||  
Db 20 GGFNNSPQFVSNNVG 35

RESULT 11  
B53895  
protegrin 4 precursor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 16-Jul-1999  
C:Accession: B53895  
R:Zhao, C.; Liu, L.; Lehrer, R.I.  
FEBS Lett. 346, 285-288, 1994  
A:Title: Identification of a new member of the protegrin family by cDNA cloning.  
A:Reference number: S45712; MUID:94283613  
A:Accession: B53895  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-149 <ZHA>  
A:CROSS-references: GB:X83268; NID:g603039; PIDN:CAA58241.1; PID:g603040  
C:Superfamily: cathelin; cystatin homology  
C:Keywords: amidated carboxyl end; antibacterial; neutrophil

F:1-29/Domain: signal sequence #status predicted <SIG>  
F:22-129/Domain: cystatin homology <CYS>  
F:131-148/Product: protegrin 4 #status predicted <MAT>  
F:148/Modified site: amidated carboxyl end (Arg) (amide in mature form from following

Query Match 48.3%; Score 43; DB 2; Length 149;  
Best Local Similarity 55.6%; Pred. No. 5.4;  
Matches 10; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 RGRLSYRRRFSVSVGR 18  
||||| : : : :  
Db 131 RGRLCYCRGWICFCVGR 148

RESULT 12  
T03240  
FLO/LFY protein homolog NFL1 - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Mar-2001  
C:Accession: T03240  
R:Kelly, A.J.; Bonlander, M.B.; Meeks-Wagner, D.R.  
Plant Cell 7, 225-234, 1995  
A:Title: NFL, the tobacco homologue of FLORICAULA and LEAFY, is transcriptionally ex  
A:Reference number: Z14855; MUID:95276463  
A:Accession: T03240  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-413 <REL>  
A:CROSS-references: EMBL:U16172; NID:g561681; PIDN:AAC48985.1; PID:g561683  
A:Experimental source: cultivar Samsun  
C:Genetics:  
A:Introns: 154/1; 288/3  
C:Superfamily: Arabidopsis thaliana LFY floral meristem identity control protein  
C:Keywords: transcription regulation

Query Match 48.3%; Score 43; DB 2; Length 413;  
Best Local Similarity 52.9%; Pred. No. 14;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGRLSYRRRFSVSVGR 18  
||| : : : || : ||  
Db 186 GGRMKQRRKKVSTGR 202

RESULT 13  
T30942  
aminopeptidase (EC 3.4.11.-) - Indian meal moth  
C:Species: Plodia interpunctella (Indian meal moth)  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
C:Accession: T30942  
R:Zhu, Y.C.; Oppert, B.; Kramer, K.J.; McGaughey, W.H.; Dowdy, A.K.  
submitted to the EMBL Data Library, November 1997  
A:Description: Molecular comparison of aminopeptidase cDNAs and gene structure between  
A:Reference number: Z20942  
A:Accession: T30942  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1016 <ZHU>  
A:CROSS-references: EMBL:AF034483; NID:g2645992; PID:g2645993; PIDN:AAC36148.1  
C:Superfamily: membrane alanyl aminopeptidase  
C:Keywords: alpha-aminoacylpeptide hydrolase

Query Match 48.3%; Score 43; DB 2; Length 1016;  
Best Local Similarity 53.3%; Pred. No. 33;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 GRLSYRRRFSVSVG 17  
||| : ||| : : :  
Db 555 GLTISQRFDITNG 569

RESULT 14  
T30943  
aminopeptidase (EC 3.4.11.-) - Indian meal moth  
C:Species: Plodia interpunctella (Indian meal moth)  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
C:Accession: T30943  
R:Zhu, Y.C.; Oppert, B.; Kramer, K.J.; McGaughey, W.H.; Dowdy, A.K.  
submitted to the EMBL Data Library, November 1997  
A:Description: Molecular comparison of aminopeptidase cDNAs and gene structure between t  
A:Reference number: Z20942  
A:Accession: T30943  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1016 <ZHU>  
A:Cross-references: EMBL:AF034484; NID:g2645994; PID:g2645995; PIDN:AAC36147.1  
C:Superfamily: membrane alanyl aminopeptidase  
C:Keywords: alpha-aminoacylpeptide hydrolase

Query Match 48.3%; Score 43; DB 2; Length 1016;  
Best Local Similarity 53.3%; Pred. No. 33;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 GRLSYSRRRFSVSG 17  
|||:|:|:|:|  
Db 555 GRLTISQRREFDITNG 569

RESULT 15  
C72642  
hypothetical protein APE0571 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: C72642  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339  
A:Accession: C72642  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-118 <KAW>  
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79539.1; PID:d1043325; PID:g510  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0571

Query Match 47.2%; Score 42; DB 2; Length 118;  
Best Local Similarity 81.8%; Pred. No. 6.4;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGRRLSYRRR 11  
|||||  
Db 21 RGRRLSSRRR 31

Search completed: February 12, 2002, 12:34:35  
Job time: 553 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:24:52 ; Search time 106.12 Seconds  
(without alignments)  
3.817 Million cell updates/sec

Title: US-09-485-571-15

Perfect score: 89

Sequence: 1 RGGRLSYRRRFSVSVGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	82.0	18	1	US-08-499-523-63
2	73	82.0	18	1	US-08-499-523-67
3	73	82.0	18	2	US-08-752-852A-230
4	73	82.0	18	4	US-09-128-345-63
5	73	82.0	18	4	US-09-128-345-67
6	71	79.8	18	1	US-08-499-523-53
7	71	79.8	18	1	US-08-499-523-58
8	71	79.8	18	4	US-09-128-345-53
9	71	79.8	18	4	US-09-128-345-58
10	70	78.7	18	1	US-08-499-523-54
11	70	78.7	18	1	US-08-499-523-59
12	70	78.7	18	4	US-09-128-345-54
13	70	78.7	18	4	US-09-128-345-59
14	69	77.5	18	1	US-08-095-769A-1
15	69	77.5	18	1	US-08-182-483A-2
16	69	77.5	18	1	US-08-182-483A-28
17	69	77.5	18	1	US-08-243-879A-1
18	69	77.5	18	1	US-08-243-879A-27
19	69	77.5	18	1	US-08-499-523-11
20	69	77.5	18	1	US-08-499-523-16
21	69	77.5	18	1	US-08-499-523-33
22	69	77.5	18	1	US-08-499-523-48
23	69	77.5	18	2	US-08-752-852A-1
24	69	77.5	18	2	US-08-752-852A-123
25	69	77.5	18	3	US-08-752-853-1
26	69	77.5	18	3	US-08-752-853-2
27	69	77.5	18	3	US-08-984-294-1

Sequence 11, Appl  
Sequence 16, Appl  
Sequence 33, Appl  
Sequence 48, Appl  
Sequence 36, Appl  
Sequence 2, Appl  
Sequence 10, Appl  
Sequence 2, Appl  
Sequence 10, Appl  
Sequence 17, Appl  
Sequence 25, Appl  
Sequence 16, Appl  
Sequence 24, Appl  
Sequence 37, Appl  
Sequence 45, Appl  
Sequence 37, Appl  
Sequence 45, Appl  
Sequence 65, Appl

ALIGNMENTS

RESULT 1  
US-08-499-523-63  
; Sequence 63, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KORYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 8, 13, 15)  
; OTHER INFORMATION: /note="X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"

US-08-499-523-63  
Query Match 82.0%; Score 73; DB 1; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.4e-05;

Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSVGR 18  
 ||||| | |||| | |||  
 Db 1 RGGRLXYRRRFXVXVGR 18

## RESULT 2

US-08-499-523-67  
 ; Sequence 67, Application US/08499523  
 ; Patent No. 5804558  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/499,523  
 FILING DATE: 07-JUL-1995  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MURASHIGE, KATE H.  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 2000-0540.24  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 887-1500  
 TELEFAX: (202) 887-0763  
 TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 67:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 18 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: group(6, 8, 13, 15)  
 ; OTHER INFORMATION: /note="X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"  
 US-08-499-523-67

Query Match 82.0%; Score 73; DB 1; Length 18;  
 Best Local Similarity 77.8%; Pred. No. 1.4e-05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSVGR 18  
 ||||| | |||| | |||  
 Db 1 RGGRLXYRRRFXVXVGR 18

## RESULT 3

US-08-752-852A-230  
 ; Sequence 230, Application US/08752852A  
 ; Patent No. 5994306  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chang, Conway  
 ; APPLICANT: Gu, Chee-Liang  
 ; APPLICANT: Chen, Jie  
 ; APPLICANT: Steinberg, Deborah

APPLICANT: Lehrer, Robert  
 APPLICANT: Harwig, Sylvia  
 TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
 NUMBER OF SEQUENCES: 242  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PENNIE & EDMONDS LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/752,852A  
 FILING DATE: 21-NOV-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:

ATTORNEY/AGENT INFORMATION:  
 NAME: Coruzzi, Laura A.  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 8067-034-999  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-790-9090  
 TELEFAX: 212-869-9741  
 TELEX: 66141  
 INFORMATION FOR SEQ ID NO: 230:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 18 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 US-08-752-852A-230

Query Match 82.0%; Score 73; DB 2; Length 18;  
 Best Local Similarity 77.8%; Pred. No. 1.4e-05;  
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSVGR 18  
 ||||| | |||| | |||  
 Db 1 RGGRLCYARRRFAVCVGR 18

## RESULT 4

US-09-128-345-63  
 ; Sequence 63, Application US/09128345  
 ; Patent No. 6159936  
 ; GENERAL INFORMATION:

APPLICANT: LEHRER, ROBERT I.  
 APPLICANT: HARWIG, SYLVIA S.L.  
 APPLICANT: KOKRYAKOV, VLADIMIR N.  
 TITLE OF INVENTION: PROTEGRINS  
 NUMBER OF SEQUENCES: 76  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PENNIE & EDMONDS LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-63
```

```
Query Match 82.0%; Score 73; DB 4; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.4e-05;
Matches 14; Conservative 0; Mismatches 4; Indels 0;
```

```
QY 1 RGGRLSYRRRFSVSGR 18
DB 1 RGGRLXYRRRFXVXGR 18
```

```
RESULT 5
US-09-128-345-67
; Sequence 67, Application US/09128345
; Patent No. 6159936
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARMIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,345
; FILING DATE: 03-AUG-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 8067-0054-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 8, 13, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-67
```

```
Query Match 82.0%; Score 73; DB 4; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.4e-05;
Matches 14; Conservative 0; Mismatches 4; Indels 0;
```

```
QY 1 FGGRLSYRRRFSVSGR 18
DB 1 FGGRLXYRRRFXVXGR 18
```

```
RESULT 6
US-08-499-523-53
; Sequence 53, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARMIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 8..13
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: group(6, 15)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
; OTHER INFORMATION: small, or a large polar amino acid"
US-08-499-523-53
```

```
Query Match 79.8%; Score 71; DB 1; Length 18;
Best Local Similarity 77.8%; Pred. No. 2.9e-05;
Matches 14; Conservative 0; Mismatches 4; Indels 0;
```

Qy 1 RGGRLSYRRRFSVSVGR 18  
 ||||| | |||| | ||||  
 Db 1 RGGRLXYCRRRFCVXVGR 18

## RESULT 7

US-08-499-523-58  
 ; Sequence 58, Application US/08499523  
 ; Patent No. 5804558  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,523  
 ; FILING DATE: 07-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.24  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEX: (202) 887-0763  
 ; TELEFAX: 90-4030  
 ; INFORMATION FOR SEQ ID NO: 58:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: Disulfide-bond  
 ; LOCATION: 6..15  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: group(8, 13)  
 ; OTHER INFORMATION: /note= "X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"  
 US-08-499-523-58

Query Match 79.8%; Score 71; DB 1; Length 18;  
 Best Local Similarity 77.8%; Pred. No. 2.9e-05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGGRLSYRRRFSVSVGR 18  
 ||||| | |||| | ||||  
 Db 1 RGGRLXYCRRRFCVXVGR 18

## RESULT 8

US-09-128-345-53  
 ; Sequence 53, Application US/09128345  
 ; Patent No. 6159936  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.

; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/128,345  
 ; FILING DATE: 03-AUG-1998  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura, A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 8067-0054-999  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 53:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: Disulfide-bond  
 ; LOCATION: 8..13  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: group(6, 15)  
 ; OTHER INFORMATION: /note= "X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"  
 US-09-128-345-53

Query Match 79.8%; Score 71; DB 4; Length 18;  
 Best Local Similarity 77.8%; Pred. No. 2.9e-05;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGGRLSYRRRFSVSVGR 18  
 ||||| | |||| | ||||  
 Db 1 RGGRLXYCRRRFCVXVGR 18

## RESULT 9

US-09-128-345-58  
 ; Sequence 58, Application US/09128345  
 ; Patent No. 6159936  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

```
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/128,345
;; FILING DATE: 03-AUG-1998
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Coruzzi, Laura, A.
;; REGISTRATION NUMBER: 30,742
;; REFERENCE/DOCKET NUMBER: 8067-0054-999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-9741
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 58:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 18 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: Disulfide-bond
;; LOCATION: 6..15
;; NAME/KEY: Modified-site
;; LOCATION: group(8, 13)
;; OTHER INFORMATION: /note= "X is a hydrophobic, a
;; OTHER INFORMATION: small, or a large polar amino acid"
US-09-128-345-58
```

```
Query Match 79.8%; Score 71; DB 4; Length 18;
Best Local Similarity 77.8%; Pred. No. 2.9e-05;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 RGGRLSYRRRFSVSGR 18
      ||||| | |||| | |||
Db 1 RGGRLCYRRRFXVCVGR 18
```

```
RESULT 10
US-08-499-523-54
; Sequence 54, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
```

```
;; INFORMATION FOR SEQ ID NO: 54:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 18 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: Disulfide-bond
;; LOCATION: 8..13
;; NAME/KEY: Modified-site
;; LOCATION: group(6, 15)
;; OTHER INFORMATION: /note= "X is a hydrophobic, a
;; OTHER INFORMATION: small, or a large polar amino acid"
US-08-499-523-54
```

```
Query Match 78.7%; Score 70; DB 1; Length 18;
Best Local Similarity 72.2%; Pred. No. 4.2e-05;
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 RGGRLSYRRRFSVSGR 18
      ||||| | |||| | |||
Db 1 RGGRLCYRRRFXVCVGR 18
```

```
RESULT 11
US-08-499-523-59
; Sequence 59, Application US/08499523
; Patent No. 5804558
; GENERAL INFORMATION:
; APPLICANT: LEHRER, ROBERT I.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: KORYAKOV, VLADIMIR N.
; TITLE OF INVENTION: PROTEGRINS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/499,523
; FILING DATE: 07-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0540.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 6..15
; NAME/KEY: Modified-site
; LOCATION: group(8, 13)
; OTHER INFORMATION: /note= "X is a hydrophobic, a
```

OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-59

Query Match 78.7%; Score 70; DB 1; Length 18;  
Best Local Similarity 72.2%; Pred. No. 4.2e-05;  
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSVGR 18  
||||| : ||| : |||  
Db 1 RGGRLCYRRRFXICVGR 18

RESULT 12  
US-09-128-345-54  
; Sequence 54, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARMIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-0054-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 54:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Disulfide-bond  
; LOCATION: 8..13  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(6, 15)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-54

Query Match 78.7%; Score 70; DB 4; Length 18;  
Best Local Similarity 72.2%; Pred. No. 4.2e-05;  
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSVGR 18  
||||| : ||| : |||  
Db 1 RGGRLCYRRRFXICVGR 18

RESULT 13  
US-09-128-345-59  
; Sequence 59, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARMIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,345  
; FILING DATE: 03-AUG-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-0054-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Disulfide-bond  
; LOCATION: 6..15  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: group(8, 13)  
; OTHER INFORMATION: /note= "X is a hydrophobic, a  
; OTHER INFORMATION: small, or a large polar amino acid"  
US-09-128-345-59

Query Match 78.7%; Score 70; DB 4; Length 18;  
Best Local Similarity 72.2%; Pred. No. 4.2e-05;  
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSVGR 18  
||||| : ||| : |||  
Db 1 RGGRLCYRRRFXICVGR 18

RESULT 14  
US-08-095-769A-1  
; Sequence 1, Application US/08095769A  
; Patent No. 5464823  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARMIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL MAMMALIAN ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500

; CITY: Washington, DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/095,769A  
; FILING DATE: 26-JUL-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 220002054020  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-095-769A-1

Query Match 77.5%; Score 69; DB 1; Length 18;  
Best Local Similarity 77.8%; Pred. No. 6e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RGGRLSYRRRRFSVSVGR 18  
      ||||| | |||| | |||  
Db 1 RGGRLCYCRRRFCVCVGR 18

RESULT 15  
US-08-182-483A-2  
; Sequence 2, Application US/08182483A  
; Patent No. 5693486  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,483A  
; FILING DATE: 13-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-182-483A-2  
  
Query Match 77.5%; Score 69; DB 1; Length 18;  
Best Local Similarity 77.8%; Pred. No. 6e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 FGGRLSYRRRRFSVSVGR 18  
      ||||| | |||| | |||  
Db 1 FGGRLCYCRRRFCVCVGR 18  
  
Search completed: February 12, 2002, 12:32:21  
Job time: 449 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:24:27 ; Search time 242.57 Seconds  
(without alignments)  
5.497 Million cell updates/sec

Title: US-09-485-571-15

Perfect score: 89

Sequence: 1 RGRRLSYRRRFSVSGR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

```

1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	18	AAW99403	Protegrin derivati
2	89	100.0	18	AAV93616	Peptide which may
3	81	91.0	18	AAW99412	Protegrin derivati
4	81	91.0	18	AAV93177	Protegrin-like pep
5	81	91.0	18	AAV93179	Protegrin-like pep
6	81	91.0	18	AAV93615	Peptide which may
7	73	82.0	18	AAW36429	Antimicrobial prot
8	73	82.0	18	AAW09084	Cationic, antimicr
9	73	82.0	18	AAW09085	Cationic, antimicr
10	72	80.9	18	AAW18151	Cationic, antimicr
11	72	80.9	18	AAW18152	Cationic, antimicr

12	70	78.7	18	AAW18153	Cationic, antimicr
13	59	77.5	18	AAV78751	Protegrin PG-1. S
14	59	77.5	18	AAV78776	Protegrin peptide
15	59	77.5	18	AAW36322	Antimicrobial prot
16	59	77.5	18	AAW36208	Antimicrobial prot
17	59	77.5	18	AAW36353	Antimicrobial prot
18	59	77.5	18	AAW35578	Antimicrobial pept
19	69	77.5	18	AAW18144	Cationic, antimicr
20	69	77.5	18	AAW18130	Cationic, antimicr
21	69	77.5	18	AAW09073	Cationic, antimicr
22	69	77.5	18	AAW29556	Porcine protegrin
23	69	77.5	18	AAW66458	Cationic peptide p
24	69	77.5	18	AAV22018	Protegrin peptide.
25	69	77.5	18	AAV93170	Protegrin peptide
26	69	77.5	18	AAV93608	Protegrin peptide
27	69	77.5	18	AAV81680	Protegrin peptide
28	69	77.5	18	AAV91757	Cationic peptide p
29	69	77.5	18	AAV91843	Antimicrobial pept
30	69	77.5	18	AAV35050	Porcine protegrin-
31	69	77.5	149	AAW25081	Antimicrobial comp
32	69	77.5	149	AAW09087	Antimicrobial prot
33	68	76.4	18	AAV78773	Protegrin peptide
34	68	76.4	18	AAV78765	Protegrin peptide
35	68	76.4	18	AAW18147	Cationic, antimicr
36	68	76.4	18	AAW18148	Cationic, antimicr
37	68	76.4	18	AAW18149	Cationic, antimicr
38	66	74.2	18	AAW18150	Cationic, antimicr
39	66	74.2	18	AAV93659	Peptide which may
40	65	73.0	18	AAW36285	Antimicrobial prot
41	65	73.0	18	AAV22005	Antimicrobial pept
42	65	73.0	18	AAV22007	Antimicrobial pept
43	64	71.9	17	AAW36276	Antimicrobial prot
44	64	71.9	17	AAW09079	Cationic, antimicr
45	64	71.9	18	AAV78774	Protegrin peptide

#### ALIGNMENTS

RESULT 1

AAW99403  
ID AAW99403 standard; peptide; 18 AA.

XX AAW99403;

DT 08-JUN-1999 (first entry)

XX Protegrin derivative peptide SM1738.

XX Linear: protegrin; peptide antibiotic; beta-sheet; secondary structure;  
KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
KW nucleus; blood-brain barrier.

XX Synthetic.

OS XX

PN WO9907728-A2.

XX 18-FEB-1999.

PD XX

PF 06-AUG-1998; 98WO-FR01757.

XX 12-AUG-1997; 97FR-0010297.

XX (SYNT-) SYNT:EM SA.

XX Calas B, Chavanleu A, Grassy G, Kaczorek M;

PI WPI; 1999-190034/16.

DR Derivatives of antibiotic peptides lacking disulfide bridges - used

XX as carriers to deliver active agents into cells

PT

XX

PS Claim 7; Page 28; 37pp; French.  
 CC This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.  
 XX  
 SQ Sequence 18 AA;

Query Match 100.0%; Score 89; DB 20; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-08; Length 18;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGLSYRRRFSVSQGR 18  
 Db 1 rgrglssrrrrfsvsqgr 18  
 |||||

## RESULT 2

AA93616  
 ID AAY93616 standard; peptide; 18 AA.

XX  
 AC AAY93616;

XX  
 DT 25-SEP-2000 (first entry)

XX  
 DE Peptide which may be linked to anticancer agents.

XX  
 KW Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 cancer.

OS Unidentified.

XX  
 PN WO200032237-A1.

XX  
 PD 08-JUN-2000.

XX  
 PF 26-NOV-1999; 99WO-FR02939.

XX  
 PR 30-NOV-1998; 98FR-0015073.

XX  
 PA (SYNT-) SYNT:EM SA.

XX  
 PI Tamsamani J, Kaczorek M, Colin De Verdier A;

XX  
 DR WPI; 2000-412166/35.

XX  
 PT New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells

XX  
 PS Disclosure; Page 8; 34pp; French.

XX  
 CC The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.

XX

SQ Sequence 18 AA;

Query Match 100.0%; Score 89; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-08; Length 18;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGLSYRRRFSVSQGR 18  
 Db 1 rgrglssrrrrfsvsqgr 18  
 |||||

## RESULT 3

AA99412  
 ID AAW99412 standard; peptide; 18 AA.

XX  
 AC AAW99412;

XX  
 DT 08-JUN-1999 (first entry)

XX  
 DE Protegrin derivative peptide SM2196.

XX  
 KW Linear; protegrin; peptide antibiotic; beta-sheet; secondary structure;  
 KW disulphide bridge; antibody; ribozyme; antitumour agent; antiviral;  
 KW anti-inflammatory; mammal; cell membrane; passive transport; cytoplasm;  
 KW nucleus; blood-brain barrier.

OS Synthetic.

XX  
 PN WO9907728-A2.

XX  
 PD 18-FEB-1999.

XX  
 PF 06-AUG-1998; 98WO-FR01757.

XX  
 PR 12-AUG-1997; 97FR-0010297.

XX  
 PA (SYNT-) SYNT:EM SA.

XX  
 PI Calas B, Chavanieu A, Grassy G, Kaczorek M;

XX  
 DR WPI; 1999-190034/16.

XX  
 PT Derivatives of antibiotic peptides lacking disulfide bridges - used  
 PT as carriers to deliver active agents into cells

XX  
 PS Claim 7; Page 28; 37pp; French.

XX  
 CC This peptide represents a linear derivative of the protegrin family of  
 CC peptide antibiotics. Protegrin antibiotics form part of the peptide  
 CC antibiotic family which contain a beta-sheet secondary structure linked  
 CC by disulphide bridges. The new derivatives are linear and lack the  
 CC disulphide bridge. The novel derivatives are used to deliver active  
 CC agents to an organism, e.g. therapeutic proteins, antibodies (or their  
 CC fragments), nucleic acid, oligonucleotides, ribozymes, antitumour agents,  
 CC antivirals and anti-inflammatory, etc. The derivatives are non-toxic  
 CC and non-lytic but can cross mammalian cell membranes rapidly by a passive  
 CC mechanism, so can deliver active agents to cytoplasm and nucleus,  
 CC including crossing the blood-brain barrier.

XX  
 SQ Sequence 18 AA;

Query Match 91.0%; Score 81; DB 20; Length 18;  
 Best Local Similarity 88.9%; Pred. No. 1.6e-06;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGRGLSYRRRFSVSQGR 18  
 Db 1 rgrglssrrrrfststgr 18  
 |||||

## RESULT 4

AAY93177  
 ID AAY93177 standard; peptide; 18 AA.  
 AC AAY93177;  
 XX  
 XX 06-DEC-2000 (first entry)  
 DT  
 DE Protegrin-like peptide antibiotic Doxo-SynBI.  
 KW Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
 KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
 KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
 KW cancer; Parkinson's disease; depression; pain; meningitis.  
 XX  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Modified-site 1  
 FT /note= "linked to doxorubicin via a succinate  
 FT (-CO-(CH2)2-CO-) linker; optionally linked  
 FT to benzylpenicillin by a glycoamide linker"  
 PN WO200032236-A1.  
 XX  
 XX 08-JUN-2000.  
 PD  
 XX 26-NOV-1999; 99WO-FR02938.  
 PF  
 XX 30-NOV-1998; 98FR-0015074.  
 PR  
 XX (SYNT-) SYNT:EM SA.  
 PA  
 XX Clair P, Kaczorek M, Tamsamani J;  
 XX WPI; 2000-422871/36.  
 DR  
 XX Use of linear peptides as vectors for active ingredients, useful for  
 PT diagnosis and treatment of central nervous system diseases, can  
 PT transport agents passively across the blood-brain barrier  
 XX  
 PS Example I; Page 13; 54pp; French.  
 XX  
 CC The invention relates to the use of linear peptides, coupled to an active  
 CC agent, to prepare a composition able to cross the blood-brain barrier  
 CC for diagnosis or treatment of disorders localised in the central nervous  
 CC system. The linear peptide preferably has the formula: (a) X1- X16;  
 CC (b); BXXBXXXXBXXXXXB; or (c) BXXBXXXXBXXXXXB, where: each of X1-X16  
 CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
 CC be Trp; each B is aa containing a side chain that includes a basic group;  
 CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
 CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
 CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
 CC Peptides able to cross the BBB include protegrins, Antennapedia,  
 CC tachyplesins, transportin, etc. Of these several families have cytolytic  
 CC effects and are termed peptide antibiotics. They fall into 3 main  
 CC categories based on their structure: (i) peptides with alpha-helices,  
 CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
 CC beta-sheets, e.g. protegrin, tachyplesins, defensins; (iii) peptides  
 CC with no major structure but containing bends due to the presence of  
 CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
 CC represents a synthetic linear peptide designed on peptides able to cross  
 CC the BBB and is conjugated to a doxorubicin molecule by a succinate  
 CC linker. The peptide may also be linked to a benzylpenicillin molecule  
 CC by a glycoamide linker.  
 CC Conjugates of the linear peptides and the active agent are particularly  
 CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
 CC Parkinson's diseases, depression, pain and meningitis, but also for  
 CC studying drug behaviour in BBB models.  
 XX Sequence 18 AA;

Query Match 91.0%; Score 81; DB 21; Length 18;  
 Best Local Similarity 88.9%; Pred. No. 1.6e-06;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 RGGRLSYRRRFSVSVGR 18  
 DB 1 RGGRLSYRRRFSVGR 18  
 |||||  
 |||||  
 RESULT 5  
 AAY93179  
 ID AAY93179 standard; peptide; 18 AA.  
 XX  
 AC AAY93179;  
 XX  
 XX 06-DEC-2000 (first entry)  
 DT  
 DE Protegrin-like peptide antibiotic Dal-SynBI.  
 KW Cytostatic; neurodegenerative; antidepressant; analgesic; antimicrobial;  
 KW blood-brain barrier; diagnostic; central nervous system; protegrin;  
 KW Antennapedia; tachyplesin; peptide antibiotic; Alzheimer's disease;  
 KW cancer; Parkinson's disease; depression; pain; meningitis; dalargin.  
 XX  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Cross-links 1  
 FT /note= "cross-links to a molecule of dalargin via  
 FT a disulphide linker"  
 PN WO200032236-A1.  
 XX  
 XX 08-JUN-2000.  
 PD  
 XX 26-NOV-1999; 99WO-FR02938.  
 PF  
 XX 30-NOV-1998; 98FR-0015074.  
 PR  
 XX (SYNT-) SYNT:EM SA.  
 PA  
 XX Clair F, Kaczorek M, Tamsamani J;  
 XX WPI; 2000-422871/36.  
 DR  
 XX Use of linear peptides as vectors for active ingredients, useful for  
 PT diagnosis and treatment of central nervous system diseases, can  
 PT transport agents passively across the blood-brain barrier  
 XX  
 PS Example II; Page 20; 54pp; French.  
 XX  
 CC The invention relates to the use of linear peptides, coupled to an active  
 CC agent, to prepare a composition able to cross the blood-brain barrier  
 CC for diagnosis or treatment of disorders localised in the central nervous  
 CC system. The linear peptide preferably has the formula: (a) X1- X16;  
 CC (b); BXXBXXXXBXXXXXB; or (c) BXXBXXXXBXXXXXB, where: each of X1-X16  
 CC are amino acids (aa), of which 6-10 of them are hydrophobic and X6 must  
 CC be Trp; each B is aa containing a side chain that includes a basic group;  
 CC and each X is an aliphatic or aromatic aa. The linear peptide may be  
 CC retro forms of (a)-(c) containing D- and/or L-form aa, or a fragment  
 CC containing at least 5, preferably at least 7 consecutive aa from (a)-(c).  
 CC Peptides able to cross the BBB include protegrins, Antennapedia,  
 CC tachyplesins, transportin, etc. Of these several families have cytolytic  
 CC effects and are termed peptide antibiotics. They fall into 3 main  
 CC categories based on their structure: (i) peptides with alpha-helices,  
 CC e.g. cecropins and maganins; (ii) peptides with disulphide bond-linked  
 CC beta-sheets, e.g. protegrin, tachyplesins, defensins; (iii) peptides  
 CC with no major structure but containing bends due to the presence of  
 CC pro residues, e.g. bactericins and PR39. The peptides of the invention  
 CC are based on the Antennapedia family peptides; (b)-peptides are based on  
 CC protegrins; and (c)-peptides are based on tachyplesins. This sequence  
 CC represents a synthetic linear peptide designed on peptides able to cross  
 CC the BBB and is conjugated to a doxorubicin molecule by a succinate  
 CC linker. The peptide may also be linked to a benzylpenicillin molecule  
 CC by a glycoamide linker.  
 CC Conjugates of the linear peptides and the active agent are particularly  
 CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
 CC Parkinson's diseases, depression, pain and meningitis, but also for  
 CC studying drug behaviour in BBB models.  
 XX Sequence 18 AA;

CC protegrins; and (c) peptides are based on tachyplesins. This sequence  
 CC represents a synthetic linear peptide designed on peptides able to cross  
 CC the BBB and is conjugated to a daltargin molecule by a disulphide linker.  
 CC Conjugates of the linear peptides and the active agent are particularly  
 CC used to treat, prevent or diagnose brain cancer, Alzheimer's or  
 CC Parkinson's diseases, depression, pain and meningitis, but also for  
 CC studying drug behaviour in BBB models.  
 SQ Sequence 18 AA;

Query Match 91.0%; Score 81; DB 21; Length 18;  
 Best Local Similarity 88.9%; Pred. No. 1.6e-06;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSGR 18  
 Db 1 RGGRLSYRRRFSVSGR 18

RESULT 6  
 AAY93615  
 ID AAY93615 standard; peptide; 18 AA.  
 XX AAY93615;  
 AC AAY93615;

XX 25-SEP-2000 (first entry)

DE Peptide which may be linked to anticancer agents.

XX Anticancer agent; cancer cell; resistance; P-glycoprotein pump;  
 KW cancer.

XX Unidentified.

XX WO200032237-A1.

XX 08-JUN-2000.

XX 26-NOV-1999; 99WO-FR02939.

XX 30-NOV-1998; 98FR-0015073.

XX (SYNT-) SYNT:PM SA.

XX Tamsamani J, Kaczorek M, Collin De Verdier A;

XX WPI; 2000-412166/35.

XX New composition useful for cancer treatment and prevention, contains  
 PT anticancer agent and peptide vector that transports agent into cells

PS Disclosure; Page 8; 34pp; French.

XX The specification describes a pharmaceutical composition, which  
 CC comprises at least one anticancer agent associated with at least one  
 CC peptide that can transport it into cancer cells and which inhibits  
 CC development of resistance to the anticancer agent. By using the  
 CC peptide as a vector for delivery of the anticancer agent, mechanisms  
 CC that cause cancer cells to become resistant to the agent, particularly  
 CC the P-glycoprotein pump, are avoided. Also, peptides are easily  
 CC produced by chemical synthesis, can be coupled easily to the agent,  
 CC cross mammalian cell membranes rapidly by a passive mechanism (no  
 CC receptors required), and are non-toxic and non-lytic. The compositions  
 CC are used to treat cancer. The present sequence represents a peptide  
 CC which may be linked to the anticancer agents of the invention.

XX Sequence 18 AA;

Query Match 91.0%; Score 81; DB 21; Length 18;  
 Best Local Similarity 88.9%; Pred. No. 1.6e-06;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSGR 18  
 Db 1 RGGRLSYRRRFSVSGR 18

RESULT 7  
 AAW36429  
 ID AAW36429 standard; peptide; 18 AA.  
 XX AAW36429;  
 AC AAW36429;

XX 13-FEB-1998 (first entry)

DE Antimicrobial protegrin peptide (229).

XX Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.

XX Synthetic.

XX Sus scrofa.

XX WO9718826-A1.

XX 29-MAY-1997.

XX 22-NOV-1996; 96WO-US18544.

XX 21-NOV-1996; 96US-0752852.

XX 22-NOV-1995; 95US-0562346.

XX 17-MAY-1996; 96US-0649811.

XX 01-AUG-1996; 96US-0690921.

XX (INTR-) INTRABIOTICS PHARM INC.

XX (REGC) UNIV CALIFORNIA.

XX Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;

XX WPI; 1997-297871/27.

XX New antimicrobial protegrin peptide(s) - having activity against  
 PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
 PT (e.g. HIV)

PS Claim 23; Page 110; 130pp; English.

XX The present sequence is an antimicrobial protegrin peptide, which  
 CC has a broad spectrum of activity against microbial targets,  
 CC including gram-positive and gram-negative bacteria, yeast, fungi,  
 CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
 CC It can be used to preserve or disinfect a variety of materials,  
 CC including medical equipment, foodstuffs, cosmetics, contact lens  
 CC solutions, medicaments or other nutrient containing materials. It  
 CC can also be used for the prophylaxis or treatment of microbial  
 CC infections or diseases in plants and animals, e.g. conjunctivitis,  
 CC keratitis, corneal ulcers, stomach ulcers associated with  
 CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
 CC sepsis, endocarditis, pneumonia and other respiratory infections.  
 CC It is biostatic or biocidal against clinically relevant pathogens  
 CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
 CC Enterococcus faecium or faecalis, penicillin resistant  
 CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
 CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5

CC to 1 mg/kg/day, by injection.

XX Sequence 18 AA;

Query Match 82.0%; Score 73; DB 18; Length 18;

Best Local Similarity 77.8%; Pred. No. 3.1e-05; Indels 0; Gaps 0;

Matches 14; Conservative 2; Mismatches 2;

QY 1 RGRLSYSRRRFSVSGR 18

DB 1 rgrlcycarrfvcvgr 18

RESULT 8  
AAW09084  
ID AAW09084 standard; peptide; 18 AA.

AC AAW09084;

DT 11-AUG-1997 (first entry)

XX Cationic, antimicrobial, virus-neutralising protegrin IB-288.

XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;

KW Candida albicans; gram-negative bacteria; STD;

KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;

KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;

KW food.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1

FT /note= "Acylated"

FT Disulfide-bond 6..15

FT /note= "results in bullet form peptide"

XX WO9637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

XX 07-JUL-1995; 95US-0499523.

XX 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX Harwig SSL, Kokryakov VN, Lehrer RI;

XX WPI; 1997-033984/03.

XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
PT useful for the treatment of microbial infection, as food  
PT preservatives and in eye care solutions

XX Claim 6; Page 65; 106pp; English.

XX The present sequence is a specifically claimed example of a peptide,  
CC recombinantly produced, corresponding to the generic formula:  
CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
CC acid or proline; A17 may be absent or a basic, neutral/polar,  
CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
CC esterified forms, all of which may contain a disulphide bond to give a  
CC cysteine bridge. Peptides of this formula are designated protegrins and  
CC are useful as anti-bacterial, anti-viral and anti-fungal agents in

CC plants and animals. The protegrins confer resistance to microbial or  
CC viral infection in plants by preventing the growth of a virus or microbe  
CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
CC are particularly useful for the treatment of sexually transmitted  
CC disease caused by microorganisms e.g. Candida albicans, HIV-1, Chlamydia  
CC trachomatis, Treponema pallidum and Neisseria gonorrhoeae. They can also  
CC be used in eye care solutions and as preservatives for food. The  
CC protegrins are more effective under physiological conditions (e.g. in  
CC the presence of serum) than certain antibiotics and are non-toxic to the  
CC cells of higher organisms.

XX Sequence 18 AA;

Query Match 82.0%; Score 73; DB 18; Length 18;

Best Local Similarity 77.8%; Pred. No. 3.1e-05;

Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGRLSYSRRRFSVSGR 18

DB 1 rgrlcycarrfvcvgr 18

RESULT 9

AAW09085

ID AAW09085 standard; peptide; 18 AA.

XX AAW09085;

DT 11-AUG-1997 (first entry)

XX Cationic, antimicrobial, virus-neutralising protegrin IB-289.

XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;

KW Candida albicans; gram-negative bacteria; STD;

KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;

KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;

KW food.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1

FT /note= "Acylated"

FT Disulfide-bond 6..15

FT /note= "results in bullet form peptide"

FT Modified-site 18

FT /note= "Amidated"

XX WO9637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

XX 07-JUL-1995; 95US-0499523.

XX 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX Harwig SSL, Kokryakov VN, Lehrer RI;

XX WPI; 1997-033984/03.

XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
PT useful for the treatment of microbial infection, as food  
PT preservatives and in eye care solutions

XX Claim 6; Page 65; 106pp; English.

XX The present sequence is a specifically claimed example of a peptide,  
CC recombinantly produced, corresponding to the generic formula:  
CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)

CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC cysteine bridge. Peptides of this formula are designated protegrins and  
 CC are useful as anti-bacterial, anti-viral and anti-fungal agents in  
 CC plants and animals. The protegrins confer resistance to microbial or  
 CC viral infection in plants by preventing the growth of a virus or microbe  
 CC and inactivate the endotoxin of gram-negative bacteria. The protegrins  
 CC are particularly useful for the treatment of sexually transmitted  
 CC disease caused by microorganisms e.g. *Candida albicans*, HIV-1, Chlamydia  
 CC trachomatis, *Treponema pallidum* and *Neisseria gonorrhoeae*. They can also  
 CC be used in eye care solutions and as preservatives for food. The  
 CC protegrins are more effective under physiological conditions (e.g. in  
 CC the presence of serum) than certain antibiotics and are non-toxic to the  
 CC cells of higher organisms.  
 XX  
 SQ Sequence 18 AA;

Query Match 82.0%; Score 73; DB 18; Length 18;  
 Best Local Similarity 77.8%; Pred. No. 3.1e-05;  
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGRRLYSRRRFSVSVGR 18  
 Db 1 rgrrlcyarrfvcvgr 18  
 ||||| :|||:| |||

RESULT 10  
 AAW18151  
 ID AAW18151 standard; peptide; 18 AA.

AC AAW18151;  
 DT 11-AUG-1997 (first entry)  
 XX  
 DE Cationic, antimicrobial, virus-neutralising protegrin PC-55.  
 XX  
 KW Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW *Candida albicans*; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
 KW *Treponema pallidum*; *Neisseria gonorrhoeae*; eye care; preservative;  
 KW food.

OS Synthetic.  
 XX  
 PN WO9637508-A1.  
 XX  
 PD 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.  
 XX  
 PR 07-JUL-1995; 95US-0499523.  
 PR 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

PI Harwig SSL, Kokryakov VN, Lehrer RI;  
 XX  
 DR WPI; 1997-033984/03.

PT Cationic, antimicrobial, virus-neutralising protegrin peptide(s)  
 PT useful for the treatment of microbial infection, as food  
 PT preservatives and in eye care solutions  
 XX  
 PS Claim 6; Page 64; 106pp; English.  
 XX

CC The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar or small amino  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC cysteine bridge. This peptide is in snake form where all the  
 CC amino acid residues are replaced by a hydrophobic, small or large polar  
 CC amino acid (e.g. alanine in this case). Peptides of this formula are  
 CC designated protegrins and are useful as anti-bacterial, anti-viral and  
 CC anti-fungal agents in plants and animals. The protegrins confer  
 CC resistance to microbial or viral infection in plants by preventing the  
 CC growth of a virus or microbe and inactivate the endotoxin of gram-  
 CC negative bacteria. The protegrins are particularly useful for the  
 CC treatment of sexually transmitted disease caused by microorganisms e.g.  
 CC *Candida albicans*, HIV-1, Chlamydia trachomatis, *Treponema pallidum* and  
 CC *Neisseria gonorrhoeae*. They can also be used in eye care solutions and  
 CC as preservatives for food. The protegrins are more effective under  
 CC physiological conditions (e.g. in the presence of serum) than certain  
 CC antibiotics and are non-toxic to the cells of higher organisms.  
 XX  
 SQ Sequence 18 AA;

Query Match 80.9%; Score 72; DB 18; Length 18;  
 Best Local Similarity 72.2%; Pred. No. 4.4e-05;  
 Matches 13; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRRLYSRRRFSVSVGR 18  
 Db 1 rgrrlawrrrfavavr 18  
 |||||:||||:||||

RESULT 11  
 AAW18152  
 ID AAW18152 standard; peptide; 18 AA.

AC AAW18152;

DT 11-AUG-1997 (first entry)

DE Cationic, antimicrobial, virus-neutralising protegrin PC-56.

XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW *Candida albicans*; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
 KW *Treponema pallidum*; *Neisseria gonorrhoeae*; eye care; preservative;  
 KW food.

OS Synthetic.

XX WO9637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

XX 07-JUL-1995; 95US-0499523.

XX 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX Harwig SSL, Kokryakov VN, Lehrer RI;

XX WPI; 1997-033984/03.

XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s)

PT useful for the treatment of microbial infection, as food  
 XX preservatives and in eye care solutions

PS Claim 6; Page 64; 106pp; English.

XX The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. This peptide is in snake form where all the  
 CC cysteine residues are replaced by a hydrophobic, small or large polar  
 CC amino acid (e.g. alanine in this case). Peptides of this formula are  
 CC designated protegrins and are useful as anti-bacterial, anti-viral and  
 CC anti-fungal agents in plants and animals. The protegrins confer  
 CC resistance to microbial or viral infection in plants by preventing the  
 CC growth of a virus or microbe and inactivate the endotoxin of gram-  
 CC negative bacteria. The protegrins are particularly useful for the  
 CC treatment of sexually transmitted disease caused by microorganisms e.g.  
 CC Candida albicans, HIV-1, Chlamydia trachomatis, Treponema pallidum and  
 CC Neisseria gonorrhoeae. They can also be used in eye care solutions and  
 CC as preservatives for food. The protegrins are more effective under  
 CC physiological conditions (e.g. in the presence of serum) than certain  
 CC antibiotics and are non-toxic to the cells of higher organisms.

XX Sequence 18 AA;

Query Match 80.9%; Score 72; DB 18; Length 18;  
 Best Local Similarity 72.2%; Pred. No. 4.4e-05;  
 Matches 13; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSGR 18  
 DB 1 Rggrlayarrwvavgr 18

RESULT 12  
 AAW18153  
 ID AAW18153 standard; peptide; 18 AA.

XX AAW18153;

XX 11-AUG-1997 (first entry)

XX Cationic, antimicrobial, virus-neutralising protegrin PC-57.

XX Antibacterial; antiviral; antifungal; antibiotic; endotoxin;  
 KW Candida albicans; gram-negative bacteria; STD;  
 KW sexually transmitted disease; HIV-1; Chlamydia trachomatis;  
 KW Treponema pallidum; Neisseria gonorrhoeae; eye care; preservative;  
 KW food.

XX Synthetic.

XX WO9637508-A1.

XX 28-NOV-1996.

XX 24-MAY-1996; 96WO-US07594.

XX 07-JUL-1995; 95US-0499523.

XX 26-MAY-1995; 95US-0451832.

XX (UYCA-) UNIV CALIFORNIA LOS ANGELES.

XX

PI Harwig SSL, Kokryakov VN, Lehrer RI;  
 XX WPI; 1997-033984/03.

XX Cationic, antimicrobial, virus-neutralising protegrin peptide(s) -  
 PT useful for the treatment of microbial infection, as food  
 PT preservatives and in eye care solutions

PS Claim 5; Page 64; 106pp; English.

XX The present sequence is a specifically claimed example of a peptide,  
 CC recombinantly produced, corresponding to the generic formula:  
 CC A1-A2-A3-A4-A5-Cys-A7-Cys-A9-A10-A11-A12-Cys-A14-Cys-A16-(A17-A18)  
 CC where A1 = a basic amino acid; A2 and A3 = a small amino acid; A4 = a  
 CC basic or small amino acid; A5, A7 and A14 = a hydrophobic amino acid;  
 CC A9, A12 and A16 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid; A10 and A11 = a basic, hydrophobic, neutral/polar or small amino  
 CC acid or proline; A17 may be absent or a basic, neutral/polar,  
 CC hydrophobic or small amino acid; and A18 may be absent or a basic,  
 CC neutral/polar, hydrophobic or small amino acid. This has a charge of at  
 CC least +3 and its N-terminal acylated and/or C-terminal amidated or  
 CC esterified forms, all of which may contain a disulphide bond to give a  
 CC cysteine bridge. This peptide is in snake form where all the  
 CC cysteine residues are replaced by a hydrophobic, small or large polar  
 CC amino acid (e.g. alanine in this case). Peptides of this formula are  
 CC designated protegrins and are useful as anti-bacterial, anti-viral and  
 CC anti-fungal agents in plants and animals. The protegrins confer  
 CC resistance to microbial or viral infection in plants by preventing the  
 CC growth of a virus or microbe and inactivate the endotoxin of gram-  
 CC negative bacteria. The protegrins are particularly useful for the  
 CC treatment of sexually transmitted disease caused by microorganisms e.g.  
 CC Candida albicans, HIV-1, Chlamydia trachomatis, Treponema pallidum and  
 CC Neisseria gonorrhoeae. They can also be used in eye care solutions and  
 CC as preservatives for food. The protegrins are more effective under  
 CC physiological conditions (e.g. in the presence of serum) than certain  
 CC antibiotics and are non-toxic to the cells of higher organisms.

XX Sequence 18 AA;

Query Match 78.7%; Score 70; DB 18; Length 18;  
 Best Local Similarity 72.2%; Pred. No. 9.3e-05;  
 Matches 13; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 RGGRLSYRRRFSVSGR 18  
 DB 1 Rggrlayarrwvavgr 18

RESULT 13  
 AAR78751  
 ID AAR78751 standard; peptide; 18 AA.

XX AAR78751;

XX 08-OCT-1995 (first entry)

XX Protegrin PG-1.

XX protegrin; antibiotic; antimicrobial; antiviral; antibacterial;  
 KW antifungal;

XX Sus scrofa.

XX Synthetic.

XX WO9503325-A.

XX 02-FEB-1995.

XX 20-JUL-1994; 94WO-US08305.

XX 20-JUL-1993; 93US-0093926.

XX 26-JUL-1993; 93US-0095769.

PR 13-JAN-1994; 94US-0182483.  
 PR 17-MAY-1994; 94US-0243879.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Harwig SSL, Kokryakov VN, Lehrer RL;  
 XX WPI; 1995-075188/10.  
 DR  
 XX Antibiotic peptide-based cpds. designated protegrin(s) - are  
 PT useful for treating and preventing viral and microbial infections  
 PT and as preservatives  
 XX  
 PS Claims 1, 9, 10; Pages 56, 59; 80pp; English.  
 XX  
 CC New peptides are disclosed which are designated "protegrins". The  
 CC peptides are useful as antibacterial, antiviral and antifungal agents in  
 CC both animals and plants. The peptides are 16-18 amino acids in length  
 CC and are characterised by four invariant Cys residues at positions 6, 8,  
 CC 13 and 15 and either (1) by a characteristic pattern of basic and  
 CC hydrophobic amino acids and/or (2) being isolatable from animal (e.g.  
 CC porcine) leukocytes; or analogues of these peptides in which 1-4 of the  
 CC Cys residues is/are replaced by hydrophobic or small amino acids. The  
 CC peptides can be produced synthetically and some can be produced  
 CC recombinantly or can be isolated and purified from their native sources.  
 CC The peptides can be modified by N-acylation and/or C-terminal amidation  
 CC or esterification, and can be in linear or cystine-bridged form. D-Amino  
 CC acid residues can be present.  
 CC The present sequence is a specifically claimed example of the  
 CC protegrins, designated PG-1. It can be synthesised; or it can be isolated  
 CC from porcine leukocytes, in which case it is in amidated and in di-  
 CC cystine bridged form. A synthetic version in which all the amino acids  
 CC have D-configuration is also claimed.  
 XX  
 SQ Sequence 18 AA;

Query Match 77.5%; Score 69; DB 16; Length 18;  
 Best Local Similarity 77.8%; Pred. No. 0.00013;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RGRGLSYRRRFSVSVGR 18  
 Db 1 rgrglcyrrrrfcvcvgr 18  
 ||||| | |||| | ||||  
 ||||| | |||| | ||||

RESULT 14  
 AAR78776  
 ID AAR78776 standard; peptide; 18 AA.  
 XX  
 AC AAR78776;  
 XX  
 DT 08-OCT-1995 (first entry)  
 XX  
 DE Protegrin peptide sequence.  
 XX  
 KW protegrin; antibiotic; antimicrobial; antiviral; antibacterial;  
 KW antifungal;  
 KW  
 OS Synthetic.  
 XX  
 FN W09503325-A.  
 XX  
 PD 02-FEB-1995.  
 XX  
 PF 20-JUL-1994; 94WO-US08305.  
 XX  
 PR 20-JUL-1993; 93US-0093926.  
 PR 26-JUL-1993; 93US-0095769.  
 PR 13-JAN-1994; 94US-0182483.  
 PR 17-MAY-1994; 94US-0243879.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.

XX  
 PI Harwig SSL, Kokryakov VN, Lehrer RL;  
 XX WPI; 1995-075188/10.  
 DR  
 XX Antibiotic peptide-based cpds. designated protegrin(s) - are  
 PT useful for treating and preventing viral and microbial infections  
 PT and as preservatives  
 XX  
 PS Disclosure; Page 19; 80pp; English.  
 XX  
 CC New peptides are disclosed which are designated "protegrins". The  
 CC peptides are useful as antibacterial, antiviral and antifungal agents in  
 CC both animals and plants. The peptides are 16-18 amino acids in length  
 CC and are characterised by four invariant Cys residues at positions 6, 8,  
 CC 13 and 15 and either (1) by a characteristic pattern of basic and  
 CC hydrophobic amino acids and/or (2) being isolatable from animal (e.g.  
 CC porcine) leukocytes; or analogues of these peptides in which 1-4 of the  
 CC Cys residues is/are replaced by hydrophobic or small amino acids. The  
 CC peptides can be produced synthetically and some can be produced  
 CC recombinantly or can be isolated and purified from their native sources.  
 CC The peptides can be modified by N-acylation and/or C-terminal amidation  
 CC or esterification, and can be in linear or cystine-bridged form. D-Amino  
 CC acid residues can be present.  
 CC The present sequence is a specific example of the protegrin  
 CC analogues in which Cys(6, 8, 13, 15) have all been replaced by Ser.  
 XX  
 SQ Sequence 18 AA;

Query Match 77.5%; Score 69; DB 16; Length 18;  
 Best Local Similarity 77.8%; Pred. No. 0.00013;  
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 RGRGLSYRRRFSVSVGR 18  
 Db 1 ratrisfrrrrfsvsvgr 18  
 | : : : : ||||| |||||  
 | : : : : ||||| |||||

RESULT 15  
 AAW36322  
 ID AAW36322 standard; peptide; 18 AA.  
 XX  
 AC AAW36322;  
 XX  
 DT 13-FEB-1998 (first entry)  
 XX  
 DE Antimicrobial protegrin peptide (122).  
 XX  
 KW Antimicrobial protegrin; broad spectrum; gram-positive; bacterium;  
 KW gram-negative; bacteria; yeast; fungus; fungi; protozoa; virus;  
 KW retrovirus; HIV; human immunodeficiency virus; preservation;  
 KW disinfection; prophylaxis; treatment; infection; disease;  
 KW conjunctivitis; keratitis; corneal ulcer; stomach ulcer; STD;  
 KW Helicobacter pylori; sexually transmitted disease; oral mucositis;  
 KW gram-negative sepsis; endocarditis; pneumonia; biocidal; biostatic;  
 KW respiratory infection; urinary tract infection; MRSA; protozoan;  
 KW vancomycin resistant Enterococcus; pathogen; multi-drug resistance;  
 KW penicillin resistant Streptococcus pneumoniae; pig; porcine;  
 KW methicillin resistant Staphylococcus aureus; systemic candidiasis.  
 XX  
 OS Synthetic.  
 OS Sus scrofa.  
 XX  
 FN W09718826-A1.  
 XX  
 PD 29-MAY-1997.  
 XX  
 PF 22-NOV-1996; 96WO-US18544.  
 XX  
 PR 21-NOV-1996; 96US-0752852.  
 PR 22-NOV-1995; 95US-0562346.  
 PR 17-MAY-1996; 96US-0649811.



PR 01-AUG-1996; 96US-0690921.  
XX (INTR-) INTRABIOTICS PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Chang CC, Chen J, Gu CL, Lehrer RI, Steinberg DA;  
XX WPI; 1997-297871/27.  
DR  
XX  
XX New antimicrobial protegrin peptide(s) - having activity against  
PT bacteria, yeast, fungi, protozoa and certain strains of viruses  
PT (e.g. HIV)  
XX  
PS Claim 23; Page 107; 130pp: English.  
XX  
CC The present sequence is an antimicrobial protegrin peptide, which  
CC has a broad spectrum of activity against microbial targets, which  
CC including gram-positive and gram-negative bacteria, yeast, fungi,  
CC protozoa and certain strains of viruses and retroviruses, e.g. HIV.  
CC It can be used to preserve or disinfect a variety of materials,  
CC including medical equipment, foodstuffs, cosmetics, contact lens  
CC solutions, medicaments or other nutrient containing materials. It  
CC can also be used for the prophylaxis or treatment of microbial  
CC infections or diseases in plants and animals, e.g. conjunctivitis,  
CC keratitis, corneal ulcers, stomach ulcers associated with  
CC Helicobacter pylori, sexually transmitted diseases, gram-negative  
CC sepsis, endocarditis, pneumonia and other respiratory infections,  
CC urinary tract infections, systemic candidiasis and oral mucositis.  
CC It is biostatic or biocidal against clinically relevant pathogens  
CC exhibiting multi-drug resistance, e.g. vancomycin resistant  
CC Enterococcus faecium or faecalis, penicillin resistant  
CC Streptococcus pneumoniae and methicillin resistant Staphylococcus  
CC aureus (MRSA). It is given at a dosage of 0.1 to 5, preferably 0.5  
CC to 1 mg/kg/day, by injection.  
XX  
SQ Sequence 18 AA;

Query Match 77.5%; Score 69; DB 18; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.00013;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 RGRRLSYRRRRFSVSGR 18  
| | | | | | | | | |  
Db 1 rgrrlcyrrrfvcvgr 18

Search completed: February 12, 2002, 12:30:29  
Job time: 362 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:32:23 ; Search time 106.12 Seconds  
(without alignments)  
2.121 Million cell updates/sec

Title: US-09-485-571-23

Perfect score: 50

Sequence: 1 RRLSYRRRF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	80.0	14	2	US-08-752-852A-65
2	40	80.0	16	2	US-08-752-852A-80
3	40	80.0	17	2	US-08-752-852A-242
4	39	78.0	18	1	US-08-182-483A-28
5	39	78.0	18	1	US-08-243-879A-27
6	39	78.0	18	1	US-08-499-523-48
7	39	78.0	18	4	US-09-128-345-48
8	37	74.0	16	1	US-08-182-483A-20
9	37	74.0	16	1	US-08-243-879A-19
10	37	74.0	16	1	US-08-499-523-40
11	37	74.0	16	1	US-08-499-523-64
12	37	74.0	16	4	US-09-128-345-64
13	37	74.0	16	4	US-09-128-345-64
14	37	74.0	18	1	US-08-499-523-63
15	37	74.0	18	1	US-08-499-523-67
16	37	74.0	18	2	US-08-752-852A-86
17	37	74.0	18	2	US-08-752-852A-230
18	37	74.0	18	4	US-09-128-345-63
19	37	74.0	18	4	US-09-128-345-67
20	36	72.0	16	1	US-08-182-483A-18
21	36	72.0	16	1	US-08-243-879A-17
22	36	72.0	16	1	US-08-499-523-38
23	36	72.0	16	4	US-09-128-345-38
24	36	72.0	18	1	US-08-499-523-53
25	36	72.0	18	1	US-08-499-523-54
26	36	72.0	18	1	US-08-499-523-58
27	36	72.0	18	1	US-08-499-523-59

Sequence 53, Appl  
Sequence 54, Appl  
Sequence 58, Appl  
Sequence 59, Appl  
Sequence 66, Appl  
Sequence 240, App  
Sequence 26, Appl  
Sequence 64, Appl  
Sequence 73, Appl  
Sequence 62, Appl  
Sequence 63, Appl  
Sequence 69, Appl  
Sequence 81, Appl  
Sequence 23, Appl  
Sequence 2, Appl  
Sequence 3, Appl  
Sequence 5, Appl  
Sequence 6, Appl

28 72.0 18 4 US-09-128-345-53  
29 36 72.0 18 4 US-09-128-345-54  
30 36 72.0 18 4 US-09-128-345-58  
31 36 72.0 18 4 US-09-128-345-59  
32 35 70.0 13 2 US-08-752-852A-66  
33 35 70.0 13 2 US-08-752-852A-240  
34 35 70.0 13 3 US-08-752-853-26  
35 35 70.0 14 2 US-08-752-852A-64  
36 35 70.0 14 2 US-08-752-852A-73  
37 35 70.0 15 2 US-08-752-852A-62  
38 35 70.0 15 2 US-08-752-852A-63  
39 35 70.0 15 2 US-08-752-852A-69  
40 35 70.0 15 2 US-08-752-852A-81  
41 35 70.0 15 3 US-08-752-853-23  
42 35 70.0 16 1 US-08-095-769A-2  
43 35 70.0 16 1 US-08-182-483A-3  
44 35 70.0 16 1 US-08-182-483A-5  
45 35 70.0 16 1 US-08-182-483A-6

#### ALIGNMENTS

RESULT 1  
US-08-752-852A-65  
; Sequence 65, Application US/08752852A  
; Patent No. 5994306  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Conway  
; APPLICANT: Gu, Chee-Liang  
; APPLICANT: Chen, Jie  
; APPLICANT: Steinberg, Deborah  
; APPLICANT: Lehrer, Robert  
; APPLICANT: Harwig, Sylvia  
; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
; NUMBER OF SEQUENCES: 242  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752.852A  
; FILING DATE: 21-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 8067-034-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-752-852A-65

Query Match 80.0%; Score 40; DB 2; Length 14;  
 Best Local Similarity 80.0%; Pred. No. 0.21;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLSYRRRF 10  
 |||||  
 Db 1 RRLCYRRRF 10

## RESULT 2

US-08-752-852A-80  
 ; Sequence 80, Application US/08752852A  
 ; Patent No. 5994306  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chang, Conway  
 ; APPLICANT: Gu, Chee-Liang  
 ; APPLICANT: Steinberg, Deborah  
 ; APPLICANT: Lehrner, Robert  
 ; APPLICANT: Harwig, Sylvia  
 ; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
 ; NUMBER OF SEQUENCES: 242  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/752,852A  
 ; FILING DATE: 21-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 8067-034-999  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-790-9090  
 ; TELEFAX: 212-869-9741  
 ; TELEX: 66141  
 ; INFORMATION FOR SEQ ID NO: 80:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 16 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 ; US-08-752-852A-80

Query Match 80.0%; Score 40; DB 2; Length 16;  
 Best Local Similarity 80.0%; Pred. No. 0.24;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLSYRRRF 10  
 |||||  
 Db 1 RRLCYRRRF 10

## RESULT 3

US-08-752-852A-242  
 ; Sequence 242, Application US/08752852A  
 ; Patent No. 5994306  
 ; GENERAL INFORMATION:

; APPLICANT: Chang, Conway  
 ; APPLICANT: Gu, Chee-Liang  
 ; APPLICANT: Steinberg, Deborah  
 ; APPLICANT: Lehrner, Robert  
 ; APPLICANT: Harwig, Sylvia  
 ; TITLE OF INVENTION: FINE-TUNED PROTEGRINS  
 ; NUMBER OF SEQUENCES: 242  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/752,852A  
 ; FILING DATE: 21-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 8067-034-999  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-790-9090  
 ; TELEFAX: 212-869-9741  
 ; TELEX: 66141  
 ; INFORMATION FOR SEQ ID NO: 242:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 17 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 ; US-08-752-852A-242

Query Match 80.0%; Score 40; DB 2; Length 17;  
 Best Local Similarity 80.0%; Pred. No. 0.26;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLSYRRRF 10  
 |||||  
 Db 2 RRLCYRRRF 11

## RESULT 4

US-08-182-483A-28  
 ; Sequence 28, Application US/08182483A  
 ; Patent No. 5693486  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/182,483A  
FILING DATE: 13-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-182-483A-28

Query Match 78.0%; Score 39; DB 1; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.41;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
|:|:|:|:|:|  
Db 4 RISFSRRRF 12

RESULT 5  
US-08-243-879A-27  
; Sequence 27, Application US/08243879A  
; Patent No. 5708145  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; TITLE OF INVENTION: A NEW PROTEGRIN  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington, DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/243,879A  
; FILING DATE: 17-MAY-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-243-879A-27

Query Match 78.0%; Score 39; DB 1; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.41;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
|:|:|:|:|:|  
Db 4 RISFSRRRF 12

RESULT 6  
US-08-499-523-48  
; Sequence 48, Application US/08499523  
; Patent No. 5804558  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,523  
; FILING DATE: 07-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0540.24  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-499-523-48

Query Match 78.0%; Score 39; DB 1; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.41;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
|:|:|:|:|:|  
Db 4 RISFSRRRF 12

RESULT 7  
US-09-128-345-48  
; Sequence 48, Application US/09128345  
; Patent No. 6159936  
; GENERAL INFORMATION:  
; APPLICANT: LEHRER, ROBERT I.  
; APPLICANT: HARWIG, SYLVIA S.L.  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; TITLE OF INVENTION: PROTEGRINS  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/128,345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-128-345-48

Query Match 78.0%; Score 39; DB 4; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.41;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
1:1:|||||  
DB 4 RLSYSRRRF 12

RESULT 8  
US-08-182-483A-20  
Sequence 20, Application US/08182483A  
Patent No. 5693486  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
APPLICANT: HARWIG, SYLVIA S.L.  
TITLE OF INVENTION: NOVEL ANTIBIOTIC PEPTIDES  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/182,483A  
FILING DATE: 13-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500

TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-182-483A-20

Query Match 74.0%; Score 37; DB 1; Length 16;  
Best Local Similarity 77.8%; Pred. No. 0.83;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
1:1:|||||  
DB 4 RLSYSRRRF 12

RESULT 9  
US-08-243-879A-19  
Sequence 19, Application US/08243879A  
Patent No. 5708145  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
TITLE OF INVENTION: A NEW PROTEGRIN  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington, DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/243,879A  
FILING DATE: 17-MAY-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-243-879A-19

Query Match 74.0%; Score 37; DB 1; Length 16;  
Best Local Similarity 77.8%; Pred. No. 0.83;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
1:1:|||||  
DB 4 RLSYSRRRF 12

RESULT 10  
US-08-499-523-40  
Sequence 40, Application US/08499523

Patent No. 5804558  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,523  
FILING DATE: 07-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-499-523-40

Query Match 74.0%; Score 37; DB 1; Length 16;  
Best Local Similarity 77.8%; Pred. No. 0.83;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 RLSYSRRRF 10  
||| |  
Db 4 RLGYRRRF 12

RESULT 11  
US-08-499-523-64  
Sequence 64, Application US/08499523  
Patent No. 5804558  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,523

FILING DATE: 07-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 8, 13, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-64

Query Match 74.0%; Score 37; DB 1; Length 16;  
Best Local Similarity 77.8%; Pred. No. 0.83;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 RLSYSRRRF 10  
||| |  
Db 4 RLXYRRRF 12

RESULT 12  
US-09-128-345-40  
Sequence 40, Application US/09128345  
Patent No. 6159936  
GENERAL INFORMATION:  
APPLICANT: LEHRER, ROBERT I.  
APPLICANT: HARWIG, SYLVIA S.L.  
APPLICANT: KOKRYAKOV, VLADIMIR N.  
TITLE OF INVENTION: PROTEGRINS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,345  
FILING DATE: 03-AUG-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura, A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 8067-0054-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-09-128-345-40

Query Match 74.0%; Score 37; DB 4; Length 16;  
 Best Local Similarity 77.8%; Pred. NO. 0.83;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
 || | ||||  
 Db 4 RLGYGRRRF 12

RESULT 13

US-09-128-345-64  
 ; Sequence 64, Application US/09128345  
 ; Patent No. 6159936

; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/128,345  
 ; FILING DATE: 03-AUG-1998  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura, A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 8067-0054-999  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 64:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 16 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: group(6, 8, 13, 15)  
 ; OTHER INFORMATION: /note= "X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"

Query Match 74.0%; Score 37; DB 4; Length 16;  
 Best Local Similarity 77.8%; Pred. NO. 0.83;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
 || | ||||  
 Db 4 RLXYXRRRF 12

RESULT 14

US-08-499-523-63  
 ; Sequence 63, Application US/08499523  
 ; Patent No. 5804558

; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible

; GENERAL INFORMATION:

; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,523  
 ; FILING DATE: 07-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.24  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030

; INFORMATION FOR SEQ ID NO: 63:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: group(6, 8, 13, 15)  
 ; OTHER INFORMATION: /note= "X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"

US-08-499-523-63

Query Match 74.0%; Score 37; DB 1; Length 18;  
 Best Local Similarity 77.8%; Pred. NO. 0.92;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
 || | ||||  
 Db 4 RLXYXRRRF 12

RESULT 15  
 US-08-499-523-67  
 ; Sequence 67, Application US/08499523  
 ; Patent No. 5804558

; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/499,523  
 ; FILING DATE: 07-JUL-1995  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2000-0540.24  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030

; INFORMATION FOR SEQ ID NO: 63:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 18 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: Modified-site  
 ; LOCATION: group(6, 8, 13, 15)  
 ; OTHER INFORMATION: /note= "X is a hydrophobic, a  
 ; OTHER INFORMATION: small, or a large polar amino acid"

US-08-499-523-63

Query Match 74.0%; Score 37; DB 1; Length 18;  
 Best Local Similarity 77.8%; Pred. NO. 0.92;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RLSYSRRRF 10  
 || | ||||  
 Db 4 RLXYXRRRF 12

RESULT 15  
 US-08-499-523-67  
 ; Sequence 67, Application US/08499523  
 ; Patent No. 5804558

; GENERAL INFORMATION:  
 ; APPLICANT: LEHRER, ROBERT I.  
 ; APPLICANT: HARWIG, SYLVIA S.L.  
 ; APPLICANT: KOKRYAKOV, VLADIMIR N.  
 ; TITLE OF INVENTION: PROTEGRINS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,523  
FILING DATE: 07-JUL-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0540.24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: group(6, 8, 13, 15)  
OTHER INFORMATION: /note= "X is a hydrophobic, a  
OTHER INFORMATION: small, or a large polar amino acid"  
US-08-499-523-67

Query Match 74.08; Score 37; DB 1; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.92;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 RLSYSRRRF 10  
|||  
Db 4 RLXYXRRRF 12

Search completed: February 12, 2002, 12:32:23  
Job time: 451 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:34:39 ; Search time 126.85 Seconds  
(without alignments)  
6.005 Million cell updates/sec

Title: US-09-485-571-23

Perfect score: 50

Sequence: 1 RRLSYRRRF 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	82.0	77	2 B40973	spermatid-specific
2	41	82.0	78	2 A40973	spermatid-specific
3	41	82.0	79	2 S56116	spermatid-specific
4	38	76.0	118	2 S56117	spermatid-specific
5	38	76.0	1108	2 D96798	hypothetical prote
6	36	72.0	1045	2 JC5795	CDEP protein - hum
7	35	70.0	113	2 J77123	ribosomal protein
8	35	70.0	147	2 JN0900	protegrin 2 precur
9	35	70.0	149	2 S57807	protegrin 1 precur
10	35	70.0	891	2 G84693	probable proline-r
11	34	68.0	249	2 B86273	hypothetical prote
12	34	68.0	257	1 JQ1873	BRI protein - toma
13	34	68.0	295	2 B54575	35k GTP-binding pr
14	34	68.0	295	2 I49117	Ras-like protein K
15	34	68.0	1090	2 T00533	probable DNA2-NAM7
16	33	66.0	79	2 PN0641	hypothetical prote
17	33	66.0	122	2 C53234	globulin-10 - maiz
18	33	66.0	177	2 E72281	hypothetical prote
19	33	66.0	264	2 JC6125	U2 small nuclear r
20	33	66.0	293	2 D81896	Neisseria meningit
21	33	66.0	303	2 A86443	probable major int
22	33	66.0	395	2 T19144	hypothetical prote
23	33	66.0	430	2 G96594	unknown protein, 5
24	33	66.0	447	2 H72516	hypothetical prote
25	33	66.0	516	2 S19252	1-aminocyclopropan
26	33	66.0	518	2 S31442	1-aminocyclopropan
27	33	66.0	629	2 B75330	probable ribosomal
28	33	66.0	2434	2 S44861	DNA topoisomerase
29	32	64.0	51	2 T10370	protein p6.9 - Org

## ALIGNMENTS

RESULT 1

B40973

spermatid-specific protein T2 precursor - common cuttlefish

N;Alternate names: arginine-rich protamine; testis-specific protein T2

C;Species: Sepia officinalis (common cuttlefish)

C;Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Sep-1998

C;Accession: B40973; S14086

R;Wouters-Tyrou, D.; Chartier-Harlin, M.C.; Martin-Ponthieu, A.; Boutillon, C.; Van

J. Biol. Chem. 266, 17388-17395, 1991

A;Title: Cuttlefish spermatid-specific protein T. Molecular characterization of two

A;Reference number: A40973; MUID:91373359

A;Accession: B40973

A;Molecule type: protein

A;Residues: 1-77 <WOU>

R;Martin-Ponthieu, A.; Wouters-Tyrou, D.; Belaiche, D.; Sautiere, P.; Schindler, P.

Eur. J. Biochem. 195, 611-619, 1991

A;Title: Cuttlefish sperm protamines. 1. Amino acid sequences of two distinct varia

A;Reference number: S14085; MUID:91153298

A;Accession: S14086

A;Molecule type: protein

A;Residues: 22-77 <MAR>

C;Superfamily: sperm histone

C;Keywords: DNA binding; nucleus; phosphoprotein; spermatogenesis

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-77/Product: protamine variant Sp2 #status experimental <MAT>

Query Match 82.0%; Score 41; DB 2; Length 77;  
Best Local Similarity 80.0%; Pred. No. 0.5;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10

| | | | |

Db 64 RRRSYRRRY 73

RESULT 2

A40973

spermatid-specific protein T1 precursor - common cuttlefish

N;Alternate names: arginine-rich protamine; testis-specific protein T1

C;Species: Sepia officinalis (common cuttlefish)

C;Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Sep-1998

C;Accession: A40973; S14085

R;Wouters-Tyrou, D.; Chartier-Harlin, M.C.; Martin-Ponthieu, A.; Boutillon, C.; Van

J. Biol. Chem. 266, 17388-17395, 1991

A;Title: Cuttlefish spermatid-specific protein T. Molecular characterization of two

A;Reference number: A40973; MUID:91373359

A;Accession: A40973

A;Molecule type: protein

A;Residues: 1-78 <WOU>

R;Martin-Ponthieu, A.; Wouters-Tyrou, D.; Belaiche, D.; Sautiere, P.; Schindler, P.

Eur. J. Biochem. 195, 611-619, 1991

protamine P1 - duc  
ribosomal protein  
transcription fact  
rfbp protein homol  
hypothetical prote  
pre-mRNA splicing  
sugar transport sy  
probable membrane  
hypothetical prote  
probable AMP deami  
hypothetical prote  
hypothetical prote  
hypothetical prote  
AcOrf-140 protein  
homeotic protein g

30 32 64.0 61 2 S39425  
31 32 64.0 103 2 T11985  
32 32 64.0 429 2 S30049  
33 32 64.0 447 2 C56146  
34 32 64.0 449 2 B69398  
35 32 64.0 494 2 A48133  
36 32 64.0 581 2 E86810  
37 32 64.0 612 2 S55084  
38 32 64.0 789 2 T38423  
39 32 64.0 924 2 T50996  
40 32 64.0 936 2 T34024  
41 32 64.0 1006 2 E96883  
42 32 64.0 1129 2 T25635  
43 32 64.0 1175 2 T25634  
44 31 62.0 60 2 E72867  
45 31 62.0 209 2 S35226

A:Title: Cuttlefish sperm protamines. 1. Amino acid sequences of two distinct variants.  
 A:Reference number: S14085; MUID:91153298  
 A:Accession: S14085  
 A:Molecule type: protein  
 A:Residues: 22-78 <MAT>  
 C:Superfamily: sperm histone  
 C:Keywords: DNA binding; nucleus; phosphoprotein; spermatogenesis  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-78/Product: protamine variant Sp1 #status experimental <MAT>

Query Match 82.0%; Score 41; DB 2; Length 78;  
 Best Local Similarity 80.0%; Pred. No. 0.5;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
 || |||||  
 Db 65 RRRYSRRRY 74

RESULT 3  
 S56116  
 spermatid-specific protein T1 - longfin squid  
 C:Species: Loligo pealeii (longfin squid)  
 C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 07-May-1999  
 C:Accession: S56116  
 R:Wouters-Tyrou, D.; Martin-Ponthieu, A.; Ledoux-Andula, N.; Kouach, M.; Jaquinod, M.; S  
 Biochem. J. 309, 529-534, 1995  
 A:Title: Squid spermiogenesis: molecular characterization of testis-specific pro-protami  
 A:Reference number: S56116; MUID:95351983  
 A:Accession: S56116  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-79 <WOU>  
 C:Superfamily: sperm histone

Query Match 82.0%; Score 41; DB 2; Length 79;  
 Best Local Similarity 80.0%; Pred. No. 0.51;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
 || |||||  
 Db 66 RRRYSRRRY 75

RESULT 4  
 S56117  
 spermatid-specific protein T2 precursor - longfin squid  
 N:Alternate names: sperm protamin SP  
 C:Species: Loligo pealeii (longfin squid)  
 C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 07-May-1999  
 C:Accession: S56117  
 R:Wouters-Tyrou, D.; Martin-Ponthieu, A.; Ledoux-Andula, N.; Kouach, M.; Jaquinod, M.; S  
 Biochem. J. 309, 529-534, 1995  
 A:Title: Squid spermiogenesis: molecular characterization of testis-specific pro-protami  
 A:Reference number: S56116; MUID:95351983  
 A:Accession: S56117  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-118 <WOU>

Query Match 76.0%; Score 38; DB 2; Length 118;  
 Best Local Similarity 88.9%; Pred. No. 2.8;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRR 9  
 || |||||  
 Db 65 RRRYSRRR 73

RESULT 5

D96798  
 hypothetical protein F22K20.5 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: D96798  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Al  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: D96798  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1108 <STO>  
 A:Cross-references: GB:AE005173; NID:g2829910; PIDN:AAC00618.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F22K20.5  
 A:Map position: 1

Query Match 76.0%; Score 38; DB 2; Length 1108;  
 Best Local Similarity 70.0%; Pred. No. 22;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
 || |||||  
 Db 1053 RRVFRSRRRF 1062

RESULT 6  
 JC5795  
 CDEP protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 24-Jan-1998 #sequence\_revision 13-Mar-1998 #text\_change 16-Jul-1999  
 C:Accession: JC5795  
 R:Koyano, Y.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.  
 Biochem. Biophys. Res. Commun. 241, 369-375, 1997  
 A:Title: Molecular cloning and characterization of CDEP, a novel human protein contr  
 nge factors.  
 A:Reference number: JC5795; MUID:98086358  
 A:Accession: JC5795  
 A:Molecule type: mRNA  
 A:Residues: 1-1045 <KOY>  
 A:Cross-references: DDBJ:AB008430  
 C:Comment: This protein is involved in the adhesion, proliferation, and differentia  
 C:Superfamily: pleckstrin repeat homology; protein 4.1 membrane-binding domain homol  
 F:1-374/Domain: ezrin-like #status predicted <EZR>  
 F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>  
 F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match 72.0%; Score 36; DB 2; Length 1045;  
 Best Local Similarity 60.0%; Pred. No. 50;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
 || |||||  
 Db 267 RKLSPRRRF 276

RESULT 7  
 S77123  
 ribosomal protein S6 - Synechocystis sp. (strain PCC 6803)  
 N:Alternate names: protein sll1767  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C;Accession: S77123  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
S;Reference number: S74322; MUID:97061201  
A;Accession: S77123  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-113 <KAN>  
A;Cross-references: EMBL:D90908; GB:AB001339; NID:gl652725; PIDN:BAAL7681.1; PID:gl65276  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Genetics:  
A;Gene: rps6  
A;Start codon: GTG  
C;Superfamily: Escherichia coli ribosomal protein S6  
C;Keywords: protein biosynthesis; ribosome

Query Match 70.0%; Score 35; DB 2; Length 113;  
Best Local Similarity 70.0%; Pred. No. 9.9;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRF 10  
|||:| |||  
Db 48 RRLAYQIRRF 57

RESULT 8  
JN0900  
protegrin 2 precursor - pig  
N;Alternate names: cathelin-like protein precursor; neutrophil peptide 3  
C;Species: *Sus scrofa domestica* (domestic pig)  
C;Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999  
C;Accession: JN0900; S36822; S34586  
R;Storici, P.; Zanetti, M.  
Biochem. Biophys. Res. Commun. 196, 1363-1368, 1993  
A;Title: A novel cDNA sequence encoding a pig leukocyte antimicrobial peptide with a cat  
A;Reference number: JN0900; MUID:94071898  
A;Accession: JN0900  
A;Molecule type: mRNA  
A;Residues: 1-147 <STO>  
A;Cross-references: GB:L24745; NID:g431435; PIDN:AAA31061.1; PID:g431436  
R;Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Egorov,  
FEBS Lett. 330, 339-342, 1993  
A;Title: Primary structure of three cationic peptides from porcine neutrophils. Sequence  
A;Reference number: S36820; MUID:93387466  
A;Accession: S36822  
A;Molecule type: protein  
A;Residues: 131-146 <MIR>  
R;Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.; Sh  
FEBS Lett. 327, 231-236, 1993  
A;Title: Protegrins: leukocyte antimicrobial peptides that combine features of corticost  
A;Reference number: S34585; MUID:93327946  
A;Accession: S34586  
A;Molecule type: protein  
A;Residues: 131-146 <KOK>  
C;Comment: This protein plays a role in the nonoxidative antimicrobial defense mechanism  
C;Superfamily: cathelin; cystatin homology  
C;Keywords: amidated carboxyl end; antibacterial; neutrophil  
F;1-29/Domain: signal sequence #status predicted <SIG>  
F;22-129/Domain: cystatin homology <CYS>  
F;131-146/Product: protegrin 2 #status experimental <MAT>  
F;146/Modified site: amidated carboxyl end (Val) (amide in mature form from following gl

Query Match 70.0%; Score 35; DB 2; Length 147;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 RLVSRRRF 10  
||| ||||

Db 134 RLCYRRRF 142

RESULT 9  
S57607  
protegrin 1 precursor - pig  
N;Alternate names: neutrophil peptide 1  
C;Species: *Sus scrofa domestica* (domestic pig)  
C;Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 16-Jul-1999  
C;Accession: S66284; S45712; S36820; S34585; S57607  
R;Zhao, C.; Ganz, T.; Lehrer, R.I.  
FEBS Lett. 368, 197-202, 1995  
A;Title: The structure of porcine protegrin genes.  
A;Reference number: S66283; MUID:95354835  
A;Accession: S66284  
A;Molecule type: DNA  
A;Residues: 1-149 <ZHA>  
A;Cross-references: EMBL:X84094; NID:g887642; PIDN:CAA58890.1; PID:g887643  
R;Zhao, C.; Liu, L.; Lehrer, R.I.  
FEBS Lett. 346, 285-288, 1994  
A;Title: Identification of a new member of the protegrin family by cDNA cloning.  
A;Reference number: S45712; MUID:94283613  
A;Accession: S45712  
A;Molecule type: mRNA  
A;Residues: 1-149 <ZH2>  
A;Cross-references: GB:X79868; NID:g603035; PIDN:CAA56251.1; PID:g603036  
R;Mirgorodskaya, O.A.; Shevchenko, A.A.; Abdalla, K.O.M.A.; Chernushevich, I.V.; Eg  
FEBS Lett. 330, 339-342, 1993  
A;Title: Primary structure of three cationic peptides from porcine neutrophils. Seq  
A;Reference number: S36820; MUID:93387466  
A;Accession: S36820  
A;Molecule type: protein  
A;Residues: 131-148 <MIR>  
R;Kokryakov, V.N.; Harwig, S.S.L.; Panyutich, E.A.; Shevchenko, A.A.; Aleshina, G.M.  
FEBS Lett. 327, 231-236, 1993  
A;Title: Protegrins: leukocyte antimicrobial peptides that combine features of cort  
A;Reference number: S34585; MUID:93327946  
A;Accession: S34585  
A;Molecule type: protein  
A;Residues: 131-148 <KOK>  
C;Genetics:  
A;Gene: NPG1  
A;Introns: 66/3; 102/3; 126/3  
C;Superfamily: cathelin; cystatin homology  
C;Keywords: amidated carboxyl end; antibacterial; neutrophil  
F;1-29/Domain: signal sequence #status predicted <SIG>  
F;22-129/Domain: cystatin homology <CYS>  
F;30-130/Domain: propeptide #status predicted <PRO>  
F;131-148/Product: protegrin 1 #status experimental <MAT>  
F;148/Modified site: amidated carboxyl end (Arg) (amide in mature form from followi

Query Match 70.0%; Score 35; DB 2; Length 149;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 FLVSRRRF 10  
||| ||||

Db 134 FLCYRRRF 142

RESULT 10  
G84693  
probable proline-rich protein [imported] - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: G84693  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallo  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A;Reference number: A84420; MUID:20083487

A:Accession: G84693  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-891 <STO>  
A:Cross-references: GB:AE002093; NID:g3980411; PIDN:AAC95214.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g29210  
A:Map position: 2

Query Match 70.0%; Score 35; DB 2; Length 891;  
Best Local Similarity 88.9%; Pred. No. 67;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRR 9  
||| |||

DB 265 RRLSNRRR 273

## RESULT 11

B86273  
hypothetical protein AAD3285.1 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: B86273  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: B86273  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-249 <STO>  
A:Cross-references: GB:AE005172; NID:g5080775; PIDN:AAD3285.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 68.0%; Score 34; DB 2; Length 249;  
Best Local Similarity 77.8%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRR 9  
||| |||

DB 49 RRLSRRR 57

## RESULT 12

JQ1873  
BRI protein - tomato mottle virus (isolate Florida)  
C:Species: tomato mottle virus  
C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 23-Jul-1999  
C:Accession: JQ1873  
R:Abouzeid, A.M.; Polston, J.E.; Hiebert, E.  
J. Gen. Virol. 73, 3225-3229, 1992  
A:Title: The nucleotide sequence of tomato mottle virus, a new geminivirus isolated from  
A:Reference number: JQ1869; MUID:93107858  
A:Accession: JQ1873  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-257 <ABO>  
A:Cross-references: GB:L14461; NID:g1200530; PIDN:AAC32418.1; PID:g295328  
C:Genetics:  
A:Map position: segment B  
C:Superfamily: tomato golden mosaic virus BRI protein

Query Match 68.0%; Score 34; DB 1; Length 257;  
Best Local Similarity 70.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLSYRRR 10  
||| |||

DB 8 RGFYSNRR 17

## RESULT 13

B54575  
35K GTP-binding protein Gem - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 02-Feb-2001  
C:Accession: B54575  
R:Maguire, J.; Santoro, T.; Jensen, P.; Siebenlist, U.; Yewdell, J.; Kelly, K.  
Science 265, 241-244, 1994  
A:Title: Gem: an induced, immediate early protein belonging to the Ras family.  
A:Reference number: A54575; MUID:94294787  
A:Accession: B54575  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual.  
A:Molecule type: mRNA  
A:Residues: 1-295 <MAG>  
A:Cross-references: GB:U10551  
C:Keywords: GTP binding; immediate-early protein; nucleotide binding; P-loop  
F:81-88/Region: nucleotide-binding motif A (P-loop)  
F:190-193/Region: GTP-binding NKXD motif

Query Match 68.0%; Score 34; DB 2; Length 295;  
Best Local Similarity 66.7%; Pred. No. 37;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRR 9  
||| |||

DB 250 RRLAYQKR 258

## RESULT 14

I49117  
Ras-like protein Kir/Gem - mouse  
N:Alternate names: kinase-inducible ras-like (kir)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 05-Nov-1999  
C:Accession: I49117  
R:Cohen, L.; Mohr, R.; Chen, Y.Y.; Huang, M.; Kato, R.; Dorin, D.; Tamanol, F.; Goga  
Proc. Natl. Acad. Sci. U.S.A. 91, 12448-12452, 1994  
A:Title: Transcriptional activation of a ras-like gene (kir) by oncogenic tyrosine k  
A:Reference number: I38745; MUID:95107974  
A:Accession: I49117  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-295 <RES>  
A:Cross-references: EMBL:U13053; NID:g563155; PIDN:AAC52145.1; PID:g563156  
C:Genetics:  
A:Gene: kir

Query Match 68.0%; Score 34; DB 2; Length 295;  
Best Local Similarity 66.7%; Pred. No. 37;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRR 9  
||| |||

DB 250 RRLAYQKR 258

## RESULT 15

T00533  
probable DNA2-NAM7 helicase family protein [imported] - Arabidopsis thaliana  
N:Alternate names: SEN1 protein homolog T20K24.14  
C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 16-Feb-2001  
C:Accession: T00533; G84572  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
submitted to the EMBL Data Library, July 1997  
A:Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.  
A:Reference number: Z14167  
A:Accession: T00533  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1090 <ROU>  
A:Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176714  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: G84572  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1090 <STO>  
A:Cross-references: GB:AE002093; NID:g3176714; PIDN:AAD12029.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: T20K24.14; At2g19120  
A:Map position: 2  
A:Introns: 250/1; 310/1; 342/1; 383/2; 427/3; 1016/3

Query Match 68.0%; Score 34; DB 2; Length 1090;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 RRLSYRRRF 10  
|.:|||.:|  
Db 209 RKVYSQRSF 218

Search completed: February 12, 2002, 12:34:39  
Job time: 557 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:39:51 ; Search time 67.2 seconds  
(without alignments)  
5.456 Million cell updates/sec

Title: US-09-485-571-23  
Perfect score: 50  
Sequence: 1 RRLSYRRRF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	82.0	77	1 PRT2_SEPOF	P80002 septia offic
2	41	82.0	78	1 PRT1_SEPOF	P80001 septia offic
3	37	74.0	59	1 HSP1_MACRU	P42142 macropus ru
4	37	74.0	60	1 HSP1_CAEFU	P42131 caenolestes
5	37	74.0	60	1 HSP1_DASVI	P42133 dasyurus vi
6	37	74.0	60	1 HSP1_MACAG	P42137 macropus ag
7	37	74.0	60	1 HSP1_MACGI	P42139 macropus gi
8	37	74.0	61	1 HSP1_ANTLA	O18745 antechinomy
9	37	74.0	61	1 HSP1_ANTSW	P42130 antechinus
10	37	74.0	61	1 HSP1_MACEU	P42138 macropus eu
11	37	74.0	61	1 HSP1_MACRG	P42141 macropus ru
12	37	74.0	61	1 HSP1_PABBI	O18768 parantechin
13	37	74.0	61	1 HSP1_SARHA	P42151 sarcophilus
14	37	74.0	61	1 HSP1_TRIVU	P42152 trichosurus
15	37	74.0	62	1 HSP1_DASRO	P42134 dasykaluta
16	37	74.0	62	1 HSP1_MURLO	P42140 murexia lon
17	37	74.0	63	1 HSP1_ANTST	P42129 antechinus
18	35	70.0	113	1 RS6_SYNV3	P73636 synecocyst
19	35	70.0	147	1 PG2_PIG	P32195 sus scrofa
20	35	70.0	149	1 PGL_PIG	P32194 sus scrofa
21	34	68.0	257	1 VBR1_TMOV	O06661 tomato mott
22	34	68.0	295	1 GEM_MOUSE	P55041 mus musculus
23	34	68.0	1062	1 YAL3_HUMAN	O9Y216 homo sapien
24	33	66.0	68	1 HSP1_PSECU	P42145 pseudochiro
25	33	66.0	264	1 U2AG_DROME	O94535 drosophila
26	33	66.0	395	1 YQ37_CAEEL	O09458 caenorhabdi
27	33	66.0	517	1 IALC_DIACA	P27486 dianthus ca
28	33	66.0	1070	1 PVDG_PLAKN	P50494 plasmodium
29	33	66.0	1254	1 YNCA_CAEEL	P34544 caenorhabdi
30	32	64.0	50	1 BVCP_NPVOP	P24648 orgyia pseu
31	32	64.0	60	1 HSP1_ORNOR	P35307 ornithorhyn
32	32	64.0	61	1 HSP1_PLAIN	P42148 planigale 1
33	32	64.0	63	1 HSP1_DROAU	P42132 dromiciops

34 32 64.0 103 1 RR6\_CYACA  
35 32 64.0 465 1 YC14\_KLEPN  
36 32 64.0 494 1 SFR4\_HUMAN  
37 32 64.0 612 1 YMY8\_YEAST  
38 31 62.0 60 1 Y140\_NPVAC  
39 31 62.0 209 1 GLOB\_TOBAC  
40 31 62.0 210 1 FBPI\_PETHY  
41 31 62.0 212 1 MAD2\_PETHY  
42 31 62.0 215 1 GLOB\_ANTMA  
43 31 62.0 227 1 DEFA\_ANTMA  
44 31 62.0 231 1 MAD1\_PETHY  
45 31 62.0 232 1 AP3\_ARATH

#### ALIGNMENTS

RESULT 1  
PRT2\_SEPOF  
ID PRT2\_SEPOF STANDARD; PRT; 77 AA.  
AC P80002;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SPERMATID-SPECIFIC PROTEIN T2 [CONTAINS: SPERM PROTAMINE SP2].  
OS Sepia officinalis (Common cuttlefish).  
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Sepioida;  
OC Sepiidae; Sepia.  
OX NCBI\_TaxID=6610;  
RN [1]  
RN SEQUENCE.  
RX MEDLINE=91373359; PubMed=1894625;  
RA Wouters-Tyrou D., Chartier-Harlin M.-C., Martin-Ponchieu A.,  
Boutillon C., van Dorsseleer A., Sautiere P.;  
RT "Cuttlefish spermatid-specific protein T. Molecular characterization  
of two variants T1 and T2, putative precursors of sperm protamine  
variants S1 and S2.";  
RL J. Biol. Chem. 266:17388-17395(1991).  
RN [2]  
RN SEQUENCE OF 22-77.  
RX MEDLINE=91153298; PubMed=1999185;  
RA Martin-Ponchieu A., Wouters-Tyrou D., Belaiche D., Sautiere P.,  
Schindler P., van Dorsseleer A.;  
RT "Cuttlefish sperm protamines. 1. Amino acid sequences of two distinct  
variants.";  
RL Eur. J. Biochem. 195:611-619(1991).  
CC -!- FUNCTION: CUTTLEFISH SPERMIOGENESIS IS CHARACTERIZED BY A DOUBLE  
NUCLEAR PROTEIN TRANSITION: HISTONES -> SPERMATID-SPECIFIC  
PROTEINS (T1/T2) -> PROTAMINES (SP1/SP2). THE PROTAMINES COMPACT  
SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.  
CC -!- DEVELOPMENTAL STAGE: SPERMIOGENESIS.  
CC -!- PFM: PHOSPHORYLATED FORM MAY BE PREDOMINANT IN T2. SP2 APPEARS TO BE  
PHOSPHORYLATED IN ELONGATED SPERMATIDS, BUT DEPHOSPHORYLATED IN  
MATURE SPERM CELLS.  
CC -!- SIMILARITY: HIGH, TO S.OFFICINALIS SPERMATID-SPECIFIC PROTEIN T1.  
DR PIR; S14086; S14086.  
DR PIR; H40973; B40973.  
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
KW Testis; DNA condensation; Nuclear protein; Phosphorylation;  
KW Multigene family.  
FT CHAIN 1 77 SPERMATID-SPECIFIC PROTEIN T2.  
FT CHAIN 22 77 SPERM PROTAMINE SP2.  
FT DOMAIN 1 21 HYDROPHOBIC.  
FT DOMAIN 22 77 ARG-RICH (HIGHLY BASIC).  
FT VARIANT 1 1 MISSING (IN T2B).  
FT SEQUENCE 77 AA; 10485 MW; 0F2C1B215292E0D7 CRC64;

Query Match 82.0%; Score 41; DB 1; Length 77;  
Best Local Similarity 80.0%; Pred. No. 0.11;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
|| |||||

Db 64 RRRYSRRY 73

RESULT 2

ID PRTL\_SEPOF STANDARD; PRT; 78 AA.

AC P80001;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE SPERMATID-SPECIFIC PROTEIN T1 [CONTAINS: SPERM PROTAMINE SP1].

OS Eukarya; Chordata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.

OC Sepiidae; Sepia.

OX NCBI\_TaxID=6610;

RN [1]

RP SEQUENCE

RX MEDLINE=91373359; PubMed=1894625;

RA Wouters-Tyrou D., Chartier-Harlin M.-C., Martin-Ponthieu A., Boutillon C., van Dorselaer A., Sautiere P.; "Cuttlefish spermatid-specific protein T. Molecular characterization of two variants T1 and T2, putative precursors of sperm protamine variants Sp1 and Sp2.";

RL J. Biol. Chem. 266:17388-17395(1991).

RN [2]

RP SEQUENCE OF 22-78.

RX MEDLINE=91153298; PubMed=1999185;

RA Martin-Ponthieu A., Wouters-Tyrou D., Belaiche D., Sautiere P., Schandler P., van Dorselaer A.; "Cuttlefish sperm protamines. 1. Amino acid sequences of two distinct variants.";

RL Eur. J. Biochem. 195:611-619(1991).

CC -!- FUNCTION: CUTTLEFISH SPERMIOGENESIS IS CHARACTERIZED BY A DOUBLE NUCLEAR PROTEIN TRANSITION: HISTONES -> SPERMATID-SPECIFIC PROTEINS (T1/T2) -> PROTAMINES (SP1/SP2). THE PROTAMINES COMPACT SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- TISSUE SPECIFICITY: TESTIS.

CC -!- DEVELOPMENTAL STAGE: SPERMIOGENESIS.

CC -!- PTM: PHOSPHORYLATION OCCURS AT DIFFERENT DEGREES. THE TRIPHOSPHORYLATED FORM MAY BE PREDOMINANT IN T1. SP1 APPEARS TO BE PHOSPHORYLATED IN ELONGATED SPERMATIDS, BUT DEPHOSPHORYLATED IN MATURE SPERM CELLS.

CC -!- SIMILARITY: HIGH, TO S.OFFICINALIS SPERMATID-SPECIFIC PROTEIN T2.

DR PIR; S14085; S14085.

DR PIR; A40973; A40973.

KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding; Testis; DNA condensation; Nuclear protein; Phosphorylation; Multigene family.

KW CHAIN 1 78 SPERMATID-SPECIFIC PROTEIN T1.

FT CHAIN 22 78 SPERM PROTAMINE SP1.

FT DOMAIN 1 21 HYDROPHOBIC.

FT DOMAIN 22 78 ARG-RICH (HIGHLY BASIC).

FT VARIANT 1 1 MISSING (IN T1B).

SQ SEQUENCE 78 AA; 10631 MW; 09FE3EDBF0DCD33 CRC64;

Query Match 82.0%; Score 41; DB 1; Length 78;  
Best Local Similarity 80.0%; Pred. No. 0.11;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
|| |||||

Db 65 RRRYSRRY 74

RESULT 3

HSPI\_MACRU STANDARD; PRT; 59 AA.

ID HSPI\_MACRU

QY 1 RRLSYRRRF 10  
|| |||||

Db 65 RRRYSRRY 74

Query Match 74.0%; Score 37; DB 1; Length 59;  
Best Local Similarity 70.0%; Pred. No. 0.52;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
|| |||||

Db 43 RRRYSRRY 52

RESULT 4

HSPI\_CAERFU STANDARD; PRT; 60 AA.

ID HSPI\_CAERFU

AC P42131;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE SPERM PROTAMINE P1.

GN PRM1.

OS Caenolestes fuliginosus.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Paucituberculata; Caenolestidae; Caenolestes.

OX NCBI\_TaxID=37696;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Sperm;

RX MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.; "Molecular phylogeny and evolution of marsupial protamine P1 genes.";

RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).

CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

P42142;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE SPERM PROTAMINE P1.

GN PRM1.

OS Macropus rufus (Red kangaroo) (Megaleia rufa).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.

OX NCBI\_TaxID=9321;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Sperm;

RX MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.; "Molecular phylogeny and evolution of marsupial protamine P1 genes.";

RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).

CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- TISSUE SPECIFICITY: TESTIS.

CC -!- FUNCTION: IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L35447; AAA74616.1; --

DR InterPro; IPR000221; Protamine\_P1.

DR Pfam; PF00260; Protamine\_P1; 1.

DR PROSITE; PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding; Testis; DNA condensation; Nuclear protein.

FT INIT\_MET 0 BY SIMILARITY.

FT SEQUENCE 59 AA; 8230 MW; 78FAE592B4B2FA2 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 59;  
Best Local Similarity 70.0%; Pred. No. 0.52;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
|| |||||

Db 43 RRRYSRRY 52

RESULT 4

HSPI\_CAERFU STANDARD; PRT; 60 AA.

ID HSPI\_CAERFU

AC P42131;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE SPERM PROTAMINE P1.

GN PRM1.

OS Caenolestes fuliginosus.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Paucituberculata; Caenolestidae; Caenolestes.

OX NCBI\_TaxID=37696;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Sperm;

RX MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.; "Molecular phylogeny and evolution of marsupial protamine P1 genes.";

RL Proc. R. Soc. Lond., B. Biol. Sci. 259:7-14(1995).

CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35332; AAA74598.1; -.
DR InterPro: IPR000221; Protamine_P1.
DR Pfam: PF00260; protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 60 AA; 8514 MW; 7630E63AD33A9B05 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 60;
Best Local Similarity 70.0%; Pred. No. 0.53;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLSYSSRRRF 10
Db 43 RRGYSRRRY 52

RESULT 5
HSP1_DASVI
ID HSP1_DASVI STANDARD; PRT; 60 AA.
AC P42135; P42133;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 36, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Dasyurus viverrinus (Southeastern quoll), and
OS Dasyurus hallucinus (Satanellus/northern quoll).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasyurinae;
OX NCBI_TaxID=9279, 9280;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SpERM;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Western M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35340; AAA74599.1; -.
DR EMBL: L35341; AAA56795.1; -.
DR InterPro: IPR000221; Protamine_P1.
DR Pfam: PF00260; protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0 0 BY SIMILARITY.

```

```

SQ SEQUENCE 60 AA; 8246 MW; 615D3D85E7123025 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 60;
Best Local Similarity 70.0%; Pred. No. 0.53;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLSYSSRRRF 10
Db 44 RRGYSRRRY 53

RESULT 6
HSP1_MACAG
ID HSP1_MACAG STANDARD; PRT; 60 AA.
AC P42137;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Macropus agilis (Agile wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9313;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SpERM;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Western M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35451; AAA74615.1; -.
DR InterPro: IPR000221; Protamine_P1.
DR Pfam: PF00260; protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 60 AA; 8338 MW; 96255C818921EB85 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 60;
Best Local Similarity 70.0%; Pred. No. 0.53;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLSYSSRRRF 10
Db 43 RRGYSRRRY 52

RESULT 7
HSP1_MACGI
ID HSP1_MACGI STANDARD; PRT; 60 AA.
AC P42139;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.

```

GN PRM1.  
OS Macropus giganteus (Eastern gray kangaroo).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.  
OX NCBI\_TaxID=9317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Sperm;  
RX MEDLINE=95215351; PubMed=7700877;  
RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes."  
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L35333; AAA74604.1; -  
DR InterPro: IPR000221; Protamine.P1.  
DR Pfam: PF00260; Protamine.P1; 1.  
DR PROSITE: PS00048; PROTAMINE.P1; 1.  
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
KW Testis; DNA condensation; Nuclear protein.  
FT INIT MET 0 0 BY SIMILARITY.  
SQ SEQUENCE 60 AA; 8415 MW; 1DC25C80C490BC90 CRC64;  
  
Query Match 74.0%; Score 37; DB 1; Length 60;  
Best Local Similarity 70.0%; Pred. No. 0.53;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 RRLSYSRRRF 10  
DB 44 RRGYSRRRY 53  
  
RESULT 8  
ID HSPL\_ANTLA STANDARD; PRT; 61 AA.  
AC Q18745;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Antechinus laniger.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinomys.  
OX NCBI\_TaxID=60701;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97446280; PubMed=9299228;  
RA Krajewski C., Blacket M., Buckley L., Westerman M.;  
RT "A multigene assessment of phylogenetic relationships within the  
RT dasyurid marsupial subfamily Sminthopsinae."  
RL Mol. Phylogenet. Evol. 8:236-248(1997).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF001587; AAB91377.1; -  
DR InterPro: IPR000221; Protamine.P1.  
DR Pfam: PF00260; Protamine.P1; 1.  
DR PROSITE: PS00048; PROTAMINE.P1; 1.  
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
KW Testis; DNA condensation; Nuclear protein.  
FT INIT MET 0 0 BY SIMILARITY.  
SQ SEQUENCE 61 AA; 8409 MW; E021567627E562B1 CRC64;  
  
Query Match 74.0%; Score 37; DB 1; Length 61;  
Best Local Similarity 70.0%; Pred. No. 0.54;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 RRLSYSRRRF 10  
DB 44 RRGYSRRRY 53  
  
RESULT 9  
ID HSPL\_ANTSW STANDARD; PRT; 61 AA.  
AC P42130; P42146;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Antechinus swainsonii, Phascosorex dorsalis,  
OS Neophascogale lorentzii (Long-clawed marsupial mouse),  
OS Dasyurus albopunctatus (Native cat),  
OS Dasyurus geoffroii (Chuditch/western quoll), and  
OS Dasyurus spartacus (Native cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.  
OX NCBI\_TaxID=9284, 9295, 32551, 32545, 63143, 32546;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=A.swainsonii, and P.dorsalis; TISSUE=Sperm;  
RX MEDLINE=95215351; PubMed=7700877;  
RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes."  
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=N.lorentzii, D.albopunctatus, D.geoffroii, and D.spartacus;  
RC Krajewski C., Young J., Buckley L., Woolley P.A., Westerman M.;  
RA "Reconstructing the taxonomic radiation of dasyurine marsupials with  
RT cytochrome b, 12S rRNA, and protamine P1 gene trees."  
RL J. Mammal. Evol. 4:217-236(1997).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L35338; AAB95429.1; -  
DR EMBL; L35339; AAA74601.1; -  
DR EMBL; AF010267; AAB69297.1; -

```
DR EMBL; AF010272; AAB69302.1; -
DR EMBL; AF010274; AAB69304.1; -
DR EMBL; AF010275; AAB69305.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
BY SIMILARITY.
SQ SEQUENCE 61 AA; 8390 MW; E021472785E71221 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 61;
Best Local Similarity 70.0%; Pred. No. 0.54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
DB 44 RRGYSRRRY 53

RESULT 10
HSP1_MACEU STANDARD; PRT; 61 AA.
AC P42138;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Macropus eugenii (Tamar wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9315;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35450; AAA74614.1; -
DR EMBL; L35451; AAA74615.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
BY SIMILARITY.
SQ SEQUENCE 61 AA; 8495 MW; 58C2925C80C49A1C CRC64;

Query Match 74.0%; Score 37; DB 1; Length 61;
Best Local Similarity 70.0%; Pred. No. 0.54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
DB 44 RRGYSRRRY 53

RESULT 10
HSP1_MACEU STANDARD; PRT; 61 AA.
AC P42138;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Macropus eugenii (Tamar wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9315;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35450; AAA74614.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
BY SIMILARITY.
SQ SEQUENCE 61 AA; 8495 MW; 58C2925C80C49A1C CRC64;

Query Match 74.0%; Score 37; DB 1; Length 61;
Best Local Similarity 70.0%; Pred. No. 0.54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
DB 44 RRGYSRRRY 53
```

```
RESULT 11
HSP1_MACRG STANDARD; PRT; 61 AA.
ID P42141; P42153;
AC P42141; P42153;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Macropus rufogriseus (Red-necked wallaby), and
OS Wallabia bicolor (Swamp wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9320, 9330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35329; AAA74610.1; -
DR EMBL; L35328; AAA74609.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
BY SIMILARITY.
SQ SEQUENCE 61 AA; 8571 MW; B822925C80C490BC CRC64;

Query Match 74.0%; Score 37; DB 1; Length 61;
Best Local Similarity 70.0%; Pred. No. 0.54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
DB 44 RRGYSRRRY 53

RESULT 12
HSP1_PARBI STANDARD; PRT; 61 AA.
ID HSP1_PARBI
AC O18768;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Parantechinus bilarni (Broad-footed marsupial mouse)..
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Parantechinus.
OX NCBI_TaxID=32555;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L35329; AAA74610.1; -
DR EMBL; L35328; AAA74609.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
BY SIMILARITY.
SQ SEQUENCE 61 AA; 8571 MW; B822925C80C490BC CRC64;

Query Match 74.0%; Score 37; DB 1; Length 61;
Best Local Similarity 70.0%; Pred. No. 0.54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
DB 44 RRGYSRRRY 53
```

FT cytochrome b, 12S rRNA, and protamine P1 gene trees.";

RL J. Mammal. Evol. 4:217-236(1997).

CC -!- FUNCTION: PROTEIN SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF

CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT

CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- TISSUE SPECIFICITY: TESTIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF010277; AAB69307.1; -

DR InterPro; IPR000221; Protamine\_P1.

DR Pfam; PF00260; protamine\_P1; 1.

DR PROSITE; PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;

KW Testis; DNA condensation; Nuclear protein.

FT INIT\_MET 0 BY SIMILARITY.

SQ SEQUENCE 61 AA; 8421 MW; C02857DF087FC9A9 CRC64;

-----

Query Match 74.0%; Score 37; DB 1; Length 61;

Best Local Similarity 70.0%; Pred. No. 0.54;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10

DB 45 RRGYSRRRY 54

-----

RESULT 13

HSPI\_SARHA STANDARD; PRT; 61 AA.

AC P42151;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE SPERM PROTAMINE P1.

GN PRM1.

OS Sarcophilus harrisi (Tasmanian Devil), and

OS Dasyurus maculatus (Tiger quoll).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sarcophilus.

OX NCBI\_TaxID=9305, 9281;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.harrisi; TISSUE=Sperm;

RX MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;

RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.;"

RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=D.maculatus;

RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;

RT "Reconstructing the taxonomic radiation of dasyurine marsupials with

RT cytochrome b, 12S rRNA, and protamine P1 gene trees.;"

RL J. Mammal. Evol. 4:217-236(1997).

CC -!- FUNCTION: PROTEIN SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF

CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT

CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- TISSUE SPECIFICITY: TESTIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L35324; AAA74608.1; -

DR EMBL; AF010276; AAB69306.1; -

DR InterPro; IPR000221; Protamine\_P1.

DR Pfam; PF00260; protamine\_P1; 1.

DR PROSITE; PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;

KW Testis; DNA condensation; Nuclear protein.

FT INIT\_MET 0 BY SIMILARITY.

SQ SEQUENCE 61 AA; 8410 MW; 4A215D3D85E71230 CRC64;

-----

Query Match 74.0%; Score 37; DB 1; Length 61;

Best Local Similarity 70.0%; Pred. No. 0.54;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10

DB 44 RRGYSRRRY 53

-----

RESULT 14

HSPI\_TRIVU STANDARD; PRT; 61 AA.

AC P42152;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE SPERM PROTAMINE P1.

GN PRM1.

OS Trichosurus vulpecula (Brush-tailed possum).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.

OX NCBI\_TaxID=9337;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Sperm;

RX MEDLINE=95215351; PubMed=7700877;

RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;

RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.;"

RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).

CC -!- FUNCTION: PROTEIN SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF

CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT

CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- TISSUE SPECIFICITY: TESTIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L32744; AAA99479.1; -

DR InterPro; IPR000221; Protamine\_P1.

DR Pfam; PF00260; protamine\_P1; 1.

DR PROSITE; PS00048; PROTAMINE\_P1; 1.

KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;

KW Testis; DNA condensation; Nuclear protein.

FT INIT\_MET 0 BY SIMILARITY.

SQ SEQUENCE 61 AA; 8571 MW; 802287E527EE816C CRC64;

-----

Query Match 74.0%; Score 37; DB 1; Length 61;

Best Local Similarity 70.0%; Pred. No. 0.54;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10

Db 44 RRGYSRRY 53

RESULT 15  
HSPI\_DASRO STANDARD; PRT; 62 AA.  
AC P42134; P42144; P42149;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Dasykaluta rosamondae, Parantechinus apicalis (Dibbler), and  
OS Pseudantechinus macdonnellensis (Fat-tailed marsupial mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasykaluta.  
OX NCBI\_TaxID=33560, 9291, 9299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Sperm;  
RX MEDLINE=95215351; PubMed=7700877;  
RA Retief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: TESTIS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L35325; AAY74605.1; -;  
DR EMBL: L35326; AAY74607.1; -;  
DR EMBL: L35337; AAY74603.1; -;  
DR InterPro: IPR000221; Protamine.P1.  
DR Pfam: PF00260; protamine.P1; 1.  
DR PROSITE: PS00048; PROTAMINE.P1; 1.  
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
KW Testis; DNA condensation; Nuclear protein.  
FT INIT-MET 0 0 BY SIMILARITY.  
SQ SEQUENCE 62 AA; 8585 MW; 99C02857DF087FC9 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 62;  
Best Local Similarity 70.0%; Pred. No. 0.54;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 RRLSYRRRF 10  
Db 45 RRGYSRRY 54

Search completed: February 12, 2002, 12:39:52  
Job time: 805 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 12:38:39 ; Search time 232.64 Seconds  
(without alignments)  
6.287 Million cell updates/sec

Title: US-09-485-571-23  
Perfect score: 50  
Sequence: 1 RRLSYSTRRRF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	76.0	256	12 Q67615	Q67615 tomato mott
2	38	76.0	1108	10 Q49281	Q49281 arabidopsis
3	37	74.0	60	6 Q9GLQ5	Q9GLQ5 peradorcas
4	37	74.0	61	6 Q9GLQ9	Q9GLQ9 mactropas pa
5	37	74.0	61	6 Q9GLQ3	Q9GLQ3 onychogalea
6	37	74.0	61	6 Q9GLP9	Q9GLP9 aepyrymnus
7	37	74.0	61	6 Q9GLP8	Q9GLP8 potorous lo
8	37	74.0	61	6 Q9GLQ1	Q9GLQ1 onychogalea
9	37	74.0	62	6 Q9TUC2	Q9TUC2 sminthopsis
10	37	74.0	62	6 Q9GLQ7	Q9GLQ7 petrogale x
11	37	74.0	62	6 Q9GLQ4	Q9GLQ4 dorcopulus
12	37	74.0	62	6 Q9GLQ0	Q9GLQ0 bettongia p
13	37	74.0	63	6 Q9TUC4	Q9TUC4 sminthopsis
14	37	74.0	63	6 Q9TUC3	Q9TUC3 sminthopsis
15	37	74.0	65	6 Q9GLQ2	Q9GLQ2 lagorcheste
16	36	72.0	256	12 Q96706	Q96706 cabbage lea
17	36	72.0	1045	4 Q9Y4F1	Q9Y4F1 homo sapien
18	36	72.0	1034	4 Q94887	Q94887 homo sapien
19	35	70.0	428	5 Q9N9B5	Q9N9B5 leishmania

20	35	70.0	477	10 Q9AVH0	Q9avh0 pisum sativ
21	35	70.0	891	10 Q9ZW08	Q9zw08 arabidopsis
22	35	70.0	894	10 Q9FVB2	Q9fwb2 arabidopsis
23	35	70.0	990	4 Q9NYF9	Q9nyf9 homo sapien
24	35	70.0	1073	10 Q9FXH1	Q9fxh1 arabidopsis
25	34	68.0	127	11 Q9Z259	Q9z259 rattus norv
26	34	68.0	226	5 Q9V5Q4	Q9v5q4 drosophila
27	34	68.0	243	2 Q9LOX6	Q9lox6 streptomyc
28	34	68.0	249	10 Q9XI88	Q9xi88 arabidopsis
29	34	68.0	316	2 Q9AP92	Q9ap92 uncultured
30	34	68.0	319	10 Q9LMF9	Q9lmf9 arabidopsis
31	34	68.0	330	5 Q9VN68	Q9vn68 drosophila
32	34	68.0	525	12 Q9DWE1	Q9dwe1 rat cytomeg
33	34	68.0	981	11 Q9ESP9	Q9esp9 mus musculu
34	34	68.0	1062	4 Q9Y2L6	Q9y2l6 homo sapien
35	34	68.0	1090	10 Q64476	Q64476 arabidopsis
36	34	68.0	1179	5 Q9NCQ0	Q9ncq0 aedes aegypt
37	34	68.0	1203	5 Q9NGZ4	Q9ngz4 drosophila
38	33	66.0	64	6 Q9GLQ1	Q9glq1 hypsiprymo
39	33	66.0	122	10 Q03863	Q03863 zea mays (m
40	33	66.0	162	2 Q9LCW4	Q9lcw4 streptomyc
41	33	66.0	177	2 Q9X0U2	Q9x0u2 thermotoga
42	33	66.0	197	5 Q9N3Q6	Q9n3q6 caenorhabdl
43	33	66.0	212	4 Q9H3L7	Q9h3l7 homo sapien
44	33	66.0	293	2 Q9J0J7	Q9juj7 neisseria m
45	33	66.0	303	10 Q9C6S9	Q9c6s9 arabidopsis

## ALIGNMENTS

RESULT	1
Q67615	PRELIMINARY;
ID	Q67615
AC	Q67615
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	TOMATO MOTTLE GEMINIVIRUS (CLONE PTFB-1) BRL (CLONE PTFB-1).
GN	BRL
OS	tomato mottle virus.
OC	viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX	NCBI_TaxID=10835;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE-93139780; PubMed-8423448;
RA	Gilbertson R.L., Hidayat S.H., Paplomatas E.J., Rojas M.R., Hou Y.M., Maxwell D.P.;
RT	"Pseudorecombination between infectious cloned DNA components of tomato mottle and bean dwarf mosaic geminiviruses.";
RL	J. Gen. Virol. 74:23-31(1993).
DR	EMBL: L02618; AAA47940.1;
DR	InterPro: IPR000263; Gemini_coat_ARI_BRL.
DR	InterPro: IPR001530; Gemini_BRL.
DR	InterPro: IPR002203; Intein.
DR	Pfam: PF01489; Gemini_BRL; 1.
DR	PRINTS: PR00223; GEMCOATARBRL.
DR	PROSITE: PS00881; PROTEIN_SPLICING; UNKNOWN.1.
SQ	SEQUENCE 256 AA; 29370 MW; 0A9A3E80AED3B692 CRC64;

Query Match 76.0%; Score 38; DB 12; Length 256;  
Best Local Similarity 80.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLSYSTRRRF 10  
| | | | |  
Db 8 RGLSYSNRRF 17

RESULT 2  
O49281  
ID O49281  
PRELIMINARY; PRT; 1108 AA.

AC 049281;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE F22K20.5 PROTEIN.  
GN F22K20.5.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,  
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,  
RA Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC002291; AAC00618.1; -  
DR Mendel; 28105; Arath; 1426; 28105.  
DR InterPro; IPR000306; Znf.FYVE.  
DR InterPro; IPR000408; RCCL.  
DR Pfam; PF00415; RCCL; 6.  
DR Pfam; PF01363; FYVE; 1.  
DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_3.  
DR PROSITE; PS0012; RCCL\_3; 7.  
DR SMART; SM00064; FYVE; 1.  
SQ SEQUENCE 1108 AA; 121276 MW; B43D6ECABD14615E CRC64;

Query Match 76.0%; Score 38; DB 10; Length 1108;  
Best Local Similarity 70.0%; Pred. No. 48;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
DB 1053 RRVFRRRF 1062  
II: :|||||

RESULT 3  
Q9GLQ5  
ID Q9GLQ5 PRELIMINARY; PRT; 60 AA.  
AC Q9GLQ5;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Peradornas concinna.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Peradornas.  
OX NCBI\_TaxID=69272;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Burk A., Springer M.S.;  
RT "The Chronicle of Kangaroo Evolution.";  
RL J. Mammal. Evol. 0:0-0(2000).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY  
CC SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).  
CC -!- SIMILARITY: TO THE PROTAMINE P1 FAMILY.  
DR EMBL; AF187538; AAG27955.1; -  
DR InterPro; IPR000221; Protamine\_P1.  
DR Pfam; PF00260; Protamine\_P1; 1.  
DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
DR Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;  
KW Nucleosome core; Spermatogenesis; Testis.  
SQ SEQUENCE 60 AA; 8436 MW; B0F0943F6F8BF58B CRC64;

Query Match 74.0%; Score 37; DB 6; Length 61;  
Best Local Similarity 70.0%; Pred. No. 4;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
DB 44 RRGYSRRY 53  
II: :|||||

RESULT 5  
Q9GLQ3  
ID Q9GLQ3 PRELIMINARY; PRT; 61 AA.  
AC Q9GLQ3;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Onychogalea fraenata (bridled nail-tailed wallaby).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Onychogalea.  
OX NCBI\_TaxID=114227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Burk A., Springer M.S.;  
RT "The Chronicle of Kangaroo Evolution.";  
RL J. Mammal. Evol. 0:0-0(2000).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT

Query Match 74.0%; Score 37; DB 6; Length 60;  
Best Local Similarity 70.0%; Pred. No. 3.9;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
DB 43 RRGYSRRY 52  
II: :|||||

RESULT 4  
Q9GLQ9  
ID Q9GLQ9 PRELIMINARY; PRT; 61 AA.  
AC Q9GLQ9;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Macropus parryi (Whiptail wallaby).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.  
OX NCBI\_TaxID=9318;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Burk A., Springer M.S.;  
RT "The Chronicle of Kangaroo Evolution.";  
RL J. Mammal. Evol. 0:0-0(2000).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY  
CC SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).  
CC -!- SIMILARITY: TO THE PROTAMINE P1 FAMILY.  
DR EMBL; AF187533; AAG27950.1; -  
DR InterPro; IPR000221; Protamine\_P1.  
DR Pfam; PF00260; Protamine\_P1; 1.  
DR PROSITE; PS00048; PROTAMINE\_P1; 1.  
DR Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;  
KW Nucleosome core; Spermatogenesis; Testis.  
SQ SEQUENCE 61 AA; 8500 MW; A07F5C81C4664B6F CRC64;

Query Match 74.0%; Score 37; DB 6; Length 61;  
Best Local Similarity 70.0%; Pred. No. 4;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
DB 44 RRGYSRRY 53  
II: :|||||

RESULT 5  
Q9GLQ3  
ID Q9GLQ3 PRELIMINARY; PRT; 61 AA.  
AC Q9GLQ3;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SPERM PROTAMINE P1.  
GN PRM1.  
OS Onychogalea fraenata (bridled nail-tailed wallaby).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Onychogalea.  
OX NCBI\_TaxID=114227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Burk A., Springer M.S.;  
RT "The Chronicle of Kangaroo Evolution.";  
RL J. Mammal. Evol. 0:0-0(2000).  
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT

```

CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -|- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL: AF187542; AAG27959.1; -.
DR InterPro: IPR000221; Protamine_P1.
DR Pfam: PF00260; Protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 61 AA; 8454 MW; CBB8EFC966E44B6A CRC64;

Query Match 74.0%; Score 37; DB 6; Length 61;
Best Local Similarity 70.0%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
   || |||||
DB 44 RRGYSRRRY 53

RESULT 6
Q9GLP9
ID Q9GLP9 PRELIMINARY; PRT; 61 AA.
AC Q9GLP9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Aepyprymnus rufescens (rufous rat-kangaroo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Aepyprymnus.
OX NCBI_TaxID=38598;
RN [1]
RP Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -|- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -|- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL: AF187547; AAG27964.1; -.
DR InterPro: IPR000221; Protamine_P1.
DR Pfam: PF00260; Protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 61 AA; 8576 MW; 7B324691290717AD CRC64;

Query Match 74.0%; Score 37; DB 6; Length 61;
Best Local Similarity 70.0%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
   || |||||
DB 45 RRGYSRRRY 54

RESULT 7
Q9GLP8
ID Q9GLP8 PRELIMINARY; PRT; 61 AA.
AC Q9GLP8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

```

```

DE PROTAMINE P1.
OS Potorous longipes (long-footed potoroo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Potorous.
OX NCBI_TaxID=55310;
RN [1]
RP Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
DR EMBL: AF187548; AAG27965.1; -.
DR InterPro: IPR000221; Protamine_P1.
DR Pfam: PF00260; Protamine_P1; 1.
SQ SEQUENCE 61 AA; 8431 MW; D745F1F638DBDCD CRC64;

Query Match 74.0%; Score 37; DB 6; Length 61;
Best Local Similarity 70.0%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
   || |||||
DB 45 RRGYSRRRY 54

RESULT 8
Q9GJO1
ID Q9GJO1 PRELIMINARY; PRT; 61 AA.
AC Q9GJO1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Onychogalea unguifera (northern nail-tailed wallaby), and
OS Dendrolagus goodfellowi (Goodfellow's tree kangaroo), and
OS Setonix brachyurus (quokka).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Onychogalea.
OX NCBI_TaxID=65626, 69260, 30670;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=O.unguifera, D.goodfellowi, and S.brachyurus;
RA Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -|- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -|- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL: AF187543; AAG27960.1; -.
DR EMBL: AF187537; AAG27954.1; -.
DR EMBL: AF187541; AAG27958.1; -.
DR InterPro: IPR000221; Protamine_P1.
DR Pfam: PF00260; Protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 61 AA; 8546 MW; F0D55C81C4664B62 CRC64;

Query Match 74.0%; Score 37; DB 6; Length 61;
Best Local Similarity 70.0%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
   || |||||
DB 44 RRGYSRRRY 53

```

```

RESULT 9
Q9TUC2 ID Q9TUC2 PRELIMINARY; PRT; 62 AA.
AC Q9TUC2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Smnthopsis longicaudata (long-tailed dunnart).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Smnthopsis.
OX NCBI_TaxID=90764;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99310778; PubMed=10381317;
RA Blacket M.J., Krajewski C., Labrinidis A., Cambron B., Cooper S.,
RA Westerman M.;
RT "Systematic relationships within the dasyurid marsupial tribe
RT Smnthopsisini--a multigene approach.";
RL Mol. Phylogenet. Evol. 12:140-155(1999).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -!- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL; AF089881; AAD5340.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 62 AA; 8640 MW; A9EEA7D7C77964A9 CRC64;

Query Match 74.0%; Score 37; DB 6; Length 62;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
Db 46 RRRGYSRRRY 55

RESULT 10
Q9GLQ7 ID Q9GLQ7 PRELIMINARY; PRT; 62 AA.
AC Q9GLQ7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PROTAMINE P1.
OS Petrogale xanthopus (Ring-tailed rock wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Petrogale.
OX NCBI_TaxID=9325;
RN [1]
RP SEQUENCE FROM N.A.
RX Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
DR EMBL; AF187535; AAG27952.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; UNKNOWN.1.
SQ SEQUENCE 62 AA; 8656 MW; 8EBE685C8089D007 CRC64;

Query Match 74.0%; Score 37; DB 6; Length 62;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
Db 46 RRRGYSRRRY 55

RESULT 11
Q9GLQ4 ID Q9GLQ4 PRELIMINARY; PRT; 62 AA.
AC Q9GLQ4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Dorcopsulus vanheurni (lesser forest wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Dorcopsulus.
OX NCBI_TaxID=69270;
RN [1]
RP SEQUENCE FROM N.A.
RX Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -!- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL; AF187539; AAG27956.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
DR Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 62 AA; 8656 MW; 3054825C8089DDB6 CRC64;

Query Match 74.0%; Score 37; DB 6; Length 62;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10
Db 45 RRRGYSRRRY 54

RESULT 12
Q9GLQ0 ID Q9GLQ0 PRELIMINARY; PRT; 62 AA.
AC Q9GLQ0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Bettongia penicillata (brush-tailed bettong).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Potoroidae; Bettongia.
OX NCBI_TaxID=69259;
RN [1]
RP SEQUENCE FROM N.A.
RX Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RL J. Mammal. Evol. 0:0-0(2000).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).

```

```

CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL; AF187546; AAG27963.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 62 AA; 8633 MW; DEB479472128DDBC CRC64;

Query Match 74.0%; Score 37; DB 6; Length 62;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRF 10
   || |||||
Db 45 RRGYSRRRY 54

RESULT 13
Q9TUC4 PRELIMINARY; PRT; 63 AA.
AC Q9TUC4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Sminthopsis bindi.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sminthopsis.
OX NCBI_TaxID=90757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93310778; PubMed=10381317;
RA Blacket M.J., Krajewski C., Labrinidis A., Cambron B., Cooper S.,
RA Westerman M.;
RT "Systematic relationships within the dasyurid marsupial tribe
RL Sminthopsini--a multigene approach.";
RM Mol. Phylogenet. Evol. 12:140-155(1999).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL; AF089873; AAD5332.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 63 AA; 8654 MW; 82A6BEA7CBB76865 CRC64;

Query Match 74.0%; Score 37; DB 6; Length 63;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRF 10
   || |||||
Db 46 RRGYSRRRY 55

RESULT 14
Q9TUC3 PRELIMINARY; PRT; 63 AA.
AC Q9TUC3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.

```

```

GN PRM1.
OS Sminthopsis griseoventer (gray-bellied dunnart).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sminthopsis.
OX NCBI_TaxID=75756;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93310778; PubMed=10381317;
RA Blacket M.J., Krajewski C., Labrinidis A., Cambron B., Cooper S.,
RA Westerman M.;
RT "Systematic relationships within the dasyurid marsupial tribe
RL Sminthopsini--a multigene approach.";
RM Mol. Phylogenet. Evol. 12:140-155(1999).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL; AF089878; AAD5337.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 63 AA; 8796 MW; 99A6A5A67BB77964 CRC64;

Query Match 74.0%; Score 37; DB 6; Length 63;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RRLYSRRRF 10
   || |||||
Db 46 RRGYSRRRY 55

RESULT 15
Q9GLQ2 PRELIMINARY; PRT; 65 AA.
AC Q9GLQ2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SPERM PROTAMINE P1.
GN PRM1.
OS Lagorcheses hirsutus (rufous hare-wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Lagorcheses.
OX NCBI_TaxID=65632;
RN [1]
RP SEQUENCE FROM N.A.
RX Burk A., Springer M.S.;
RT "The Chronicle of Kangaroo Evolution.";
RM J. Mammal. Evol. 0:0-0(2000).
CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- SIMILARITY: TO THE PROTAMINE P1 FAMILY.
DR EMBL; AF187544; AAG27961.1; -.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; DNA condensation; DNA-binding; Nuclear protein;
KW Nucleosome core; Spermatogenesis; Testis.
SQ SEQUENCE 65 AA; 9052 MW; 2D18085BB29D8A0E CRC64;

Query Match 74.0%; Score 37; DB 6; Length 65;

```

Best Local Similarity 70.0%; Pred. No. 4.3;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RRLSYRRRF 10  
   || |||||  
Db 43 RRGYSRRRY 52

Search completed: February 12, 2002, 12:38:40  
Job time: 753 sec